INTERACTIVE ANALOGICAL RETRIEVAL: PRACTICE, THEORY AND TECHNOLOGY

A Dissertation
Presented to
The Academic Faculty

by

Swaroop S. Vattam

In Partial Fulfillment
of the Requirements for the Degree
Ph.D. in Computer Science in the
College of Computing

Georgia Institute of Technology
December 2012

INTERCTIVE ANALOGICAL RETRIEVAL: PRACTICE, THEORY AND TECHNOLOGY

Approved by:

Dr. Ashok K. Goel, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Jeannette Yen
School of Biology
Georgia Institute of Technology

Dr. Janet L. Kolodner
School of Interactive Computing
Georgia Institute of Technology

Dr. Mary Lou Maher
Dept. of Software & Information Systems
University of North Carolina at Charlotte

Dr. Nancy J. Nersessian
School of Interactive Computing
Georgia Institute of Technology

Date Approved: August 22, 2012

ACKNOWLEDGEMENTS

I am very grateful to my advisor, Ashok Goel. This thesis would not have been possible without his continuous intellectual encouragement, inspiration and guidance. He has helped me in every step of the way with his insight, his questioning, his ideas, and his vision. He has been a true mentor to me and helped my get through a number of crises in my graduate career, for which I am truly grateful. I appreciate his confidence in me even at times when I doubted myself. I thank him for his comments on several drafts of this document – he has been very patient in going through my writing again and again, and helped me improve my presentation. I sincerely thank his academic, financial, emotional, and personal support during my career.

I also thank my other committee members, Professors Janet Kolodner, Nancy Nersessian, Mary Lou Maher, and Jeannette Yen who have provided constructive criticism and insightful suggestions during this work. Each of them has brought a different perspective to my work, which helped it become better. Their questions from different angles have helped me think through issues clearly and clarify my ideas. They have found time to read and comment my work even on short notice. I really appreciate their understanding, support and encouragement during this work. I like to specially thank Janet for her detailed comments on writing style, organization, and presentation. Her suggestions have helped me improve my thesis tremendously.

I am also indebted to Janet for another reason. As a former advisor, she supported me during the first two years of my career at Georgia Tech. She welcomed me into her

lab, helped me find my footing in academic research and allowed me the chance to explore and discover my true research interests.

I have greatly benefited from my discussions with several other faculty members and research staff at Georgia Tech. I would like to thank Spencer Rugaber for his feedback on numerous occasions. I would also like to thank David Majerich for helping me to analyze the data from my studies.

This research has also greatly benefitted from many discussions with other members of the Design & Intelligence Lab. I would like to thank Michael Helms, Bryan Wiltgen, Maithilee Kunda, David Joyner, and Keith McGreggor for their feedback at various stages of my work. My special gratitude to Michanel Helms and Bryan Wiltgen with whom I have shared most of my graduate life in the BID project. Without their talents, gifts, and commitment, I would not have made it this far. Without their company, it would not have been fun and enjoyable either. I thank them for encouraging me, and walking alongside me through this period. I also thank Alice Shin, my undergraduate mentee for her help with coding and modeling effort.

I would like to thank faculty and students associated with the Center for Biologically inspired Design (CBID) at Georgia Tech. I like to especially thank Professors Jeannette Yen and Mark Weissburg for welcoming me into their classroom for observation and participation. Without that opportunity, this research would not have existed. I would also like to thank the students in the ME/ISyE/MSE/PTFe/BIOL 4803 course for participating in my studies.

During the course of this work, I was supported by research grants from NSF (grant# 0855916, grant# 36466J9).

I would like to thank all my friends in and out of school who gave me constant moral support. I would like to specially mention Amit Jain and Archan Bhattacharya for being caring friends. I also want to thank my friend and former mentor Vivek Balaraman for his guidance and support when I joined his group as a rank novice in my earlier life as a researcher at TRDDC.

Last but not least, I like to acknowledge my family. I like to thank my wife and best friend, Reena, for embarking with me on this adventure. I thank her for her sacrifice and putting up with the eccentricities of a student husband. I am grateful for her invaluable help, encouragement, moral support, and cooperation at times when it was most needed. This thesis would not have been possible without her patience and companionship. I also thank my sons Upamanyu and Angshuman, who have given me much love and happiness and contributed immensely to dealing with the stresses of everyday PhD life. Finally, I also want to thank my parents for their love, support and commitment to providing me with quality educational opportunities early in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Analogy is ubiquitous in human cognition. One of the important questions related to understanding the situated nature of analogy-making is how people retrieve source analogues via their interactions with external environments. This dissertation studies interactive analogical retrieval in the context of biologically inspired design (BID). BID involves creative use of analogies to biological systems to develop solutions for complex design problems (e.g., designing a device for acquiring water in desert environments based on the analogous fog-harvesting capability of Namibian beetle). Finding the right biological analogues is one of the critical first steps in BID. Designers routinely search in online information environments in order to find their biological sources of inspiration. But this task of online bio-inspiration seeking represents an instance of interactive analogical retrieval that is extremely time consuming and challenging to accomplish.

Through a series of field studies, this dissertation uncovered the salient characteristics and challenges of online bio-inspiration seeking, including (i) low rate of encountering relevant information resources, (ii) high rate of recognition errors, and (iii) significant difficulty in comprehending information resources. An information-processing theoretic model of interactive analogical retrieval was developed in order to explain those challenges and to identify the underlying causes. Three hypotheses were proposed to ameliorate those challenges by targeting the identified causes, which respectively included (i) semantically indexing and retrieving information resources using functional models, (ii) enhancing proximal cues using visual overviews derived from functional models, and (iii) augmenting information resources with functional models as external representations to scaffold the process of understanding.

These hypothesized measures were implemented in an online information-seeking technology called *Biologue*, designed to specifically support the task of online bio-inspiration seeking. The validity of the proposed measures was investigated through a

series of experimental studies using Biologue and a deployment study. The trends from these studies are encouraging and suggest that the hypothesized measures has the potential to change the dynamics of online bio-inspiration seeking in favor of ameliorating the challenges of online bio-inspiration seeking.

My thesis thus makes four significant contributions: (1) the first field studies of online bio-inspiration seeking activity, (2) the Interactive Analogical Retrieval model that explains the causal mechanisms underlying this activity, (3) a set of theory-based guidelines for addressing the challenges of this activity, and (4) Biologue, an online information environment for supporting this activity.

CHAPTER 1

INTRODUCTION

The Domain

Biologically inspired design (Benyus, 1997; French, 1994; Vincent & Mann, 2002; Yen & Weissburg, 2007) is one of the important emerging movements in engineering design. The paradigm espouses use of analogies to biological systems in generating conceptual designs for new technological innovations. This paradigm has inspired many designers in the history of design, such as Leonardo Da Vinci, the Wright brothers, etc. But it is only over the last generation or so that the paradigm has become a movement, fueled by a growing need for environmentally sustainable design on the one hand, and driven by the desire for design creativity and innovation on the other. Some examples of important innovations emerging from this paradigm include Velcro (inspired by the attachment mechanism of burr seeds), hearing aids with enhanced directional hearing (inspired by fly's auditory system), drag-reducing surfaces (inspired by shark skin), dry adhesives (inspired by attachment mechanism of gecko feet), self-cleaning surface coatings (inspired by lotus leaf), next generation wind turbine technology (inspired by the structure of flippers of humpback whales), etc.

Target Research Population

The practice of biologically inspired design remains largely *ad hoc* with no well-established communities of practice. Accepted methodologies, best practices, or tools for systematic transfer of knowledge from biology to engineering are currently lacking. Consequently, the flow of ideas, concepts, principles, etc. from biology to engineering is mostly incidental or solution-driven. Incidental here means that the origin of the biological source of inspiration is either serendipitous or happens through *ad hoc* associations between people. Solution-driven implies that the problem-solving process

goes from solutions to problems rather than other way around: it begins with a biological source and looks for human problems to apply this solution to.

At the same time, for numerous reasons discussed in Chapter 2, more and more engineers are taking an interest in biologically inspired design as this paradigm is gaining traction in the engineering community. One implication then is that engineers working on design challenges are likely to proactively look for biological sources of inspiration rather than start with a source or wait for accidental encounters with biology, which shifts the emphasis from solution-driven to problem-driven biologically inspired design. But although engineers may be experts in their respective domains, they are likely to be novices in the domain of biological systems. In order to promote biologically inspired design, the needs of designers coming from engineering have to be better understood and fulfilled. My research specifically targets the needs of this class of novice biologically inspired designers coming from an engineering background. Henceforth, whenever I refer to designers or engineers, it is to be understood that I am referring to this particular class of biologically inspired designers.

The Task of Interest

Biologically inspired design is a complex activity that encompasses many tasks and sub-tasks. However, the focus of my research in this dissertation is limited to one of the key initial tasks of biologically inspired designing, namely *bio-inspiration seeking*. Given a target design problem, this task involves finding relevant biological systems to emulate or base the design solution upon. A biological system is considered relevant if the application of the knowledge of its workings can lead to a potentially *novel* and *useful* solution to the target design problem. The importance of this task for biologically inspired design cannot be overstated - the design outcomes in this context is largely influenced by the biological systems that are discovered as sources of inspiration through the execution of this task.

The Problem

There are an estimated 5 to 15 million species of biological organisms. If one takes into account different levels of organization of biological systems like cellular-, organ-, and ecosystem-levels, then this estimated number of biological systems increases by an order of magnitude or more. Furthermore, novice bio-inspired designers coming from engineering are not familiar with the extent, scope, and richness of biology. They may be aware of only a small fraction of this vast space of biological systems that can be drawn upon in order to develop their design solutions. The near limitless availability of biological systems to draw upon coupled with designers' lack of knowledge of this vast domain of biological systems makes bio-inspiration seeking an intellectually challenging task.

How then are designers situated in one domain (engineering) supposed to find relevant systems from the vast space of available systems that belong to a completely different and mostly unfamiliar domain (biology)? The emergence of World Wide Web has made online information seeking a daily activity for most people. Whether at work, school, or play, people have come to expect instant access to information on any topic at any place and time. This expectation carries over to the task of seeking bio-inspiration as well. Studies presented in Chapters 4 and 5 show that it is a common practice among designers to search online in order to find their biological sources if inspiration. However, those studies also indicate that the online information environments on which designers rely upon do not adequately support the task of online bio-inspiration seeking. Therefore, in spite of having online access to vast amounts of biological information, designers often struggle to find their biological sources of inspiration using the online approach. The reliance on online information environments coupled with the lack of adequate support in those environments makes an intellectually challenging task even more difficult to achieve.

Research Methodology

The overall goal of this dissertation is to better understand and support the task of online bio-inspiration seeking. The methodology adopted here in order to achieve that goal consists of four parts. First, gain familiarity with the fundamental challenges of online bio-inspiration seeking by studying the phenomenon *in situ*. Second, develop a theory of online bio-inspiration seeking that can be used to provide causal explanations underlying the challenges of online bio-inspiration seeking. Third, propose theory-based hypotheses about measures that can be taken to ameliorate those challenges, followed by an implement an information-seeking environment that implements those measures. Fourth, follow-up the implementation effort with evaluation effort by conducting studies to assess effectiveness of the proposed measures.

The methodology adopted here is influenced by the in vivo/in vitro approach proposed by Dunbar (Dunbar & Blanchette 2001) wherein the in vivo research sheds light on the "real" phenomenon and also drives the hypotheses generation, while the hypotheses themselves are evaluated through in vitro research.

This methodology provides an end-to-end approach for both understanding and supporting a real-world intellectual practice, and for bridging theory development with technology design.

Research Questions and the Core of the Dissertation

- ***RQ1: What are the fundamental challenges of online bio-inspiration seeking?***

In order to address RQ1, I conducted two *in situ* studies of biologically inspired design. Both studies were conducted in the context of ME/ISyE/MSE/PTFe/BIOL 4803, a project-based introductory course on biologically inspired design offered every year at Georgia Tech. Through these studies I uncovered the characteristics and the specific nature of challenges associated with the online bio-inspiration seeking task. Details of these studies and their findings are discussed in Chapters 4 and 5.

- ***RQ2: What are the causes underlying the challenges of online bio-inspiration seeking?***

In order to answer RQ2, it is important to gain an in-depth understanding of the nature of the online bio-inspiration seeking phenomenon. I developed a theoretical account of the online bio-inspiration seeking phenomenon which guides our understanding of this phenomenon. This theory provides two kinds of accounts of online bio-inspiration seeking. First, the *kinematics* account provides a purely descriptive account of the phenomenon based on my analysis of the bio-inspiration seeking practice in the two *in situ* studies. Second, the *dynamics* account provides an explanatory account of the phenomenon in terms of its underlying causal processes or 'mechanisms.' I developed a new information processing model called *Interactive Analogical Retrieval (IAR)* in order to explain the dynamics of online bio-inspiration seeking. Details about this theoretical work are discussed in Chapter 6.

The IAR model can be used to reason forwards from deliberate changes in the information environment to its observable effects on the online bio-inspiration seeking process of designers, or backwards from observed bio-inspiration seeking effects to the factors in the information environment causing those effects. Reasoning backwards, I provided the causal explanations for the observed challenges associated with the online bio-inspiration seeking process. These detailed explanations are discussed in Chapter 7.

Once I identified the causes, I began to address the question of how to address the challenges by targeting those causes.

- ***RQ3: What measures can be taken to ameliorate the challenges of online bio-inspiration seeking?***

The IAR model also helps us make predictions about the kind of support that can be provided to designers in online information environments in order to address the challenges that they face. These predictions become hypotheses that can be evaluated systematically. I proposed three hypotheses in this dissertation for ameliorating the

identified challenges associated with online bio-inspiration seeking. Details about these hypotheses and the rationale behind them are discussed in Chapter 8.

Based on the normative implications of the proposed hypotheses, I developed an online information-seeking environment called *Biologue* that is intended to better support the needs of designers engaged in online bio-inspiration seeking. Details about the design, architecture, and implementation of Biologue are presented in Chapter 9.

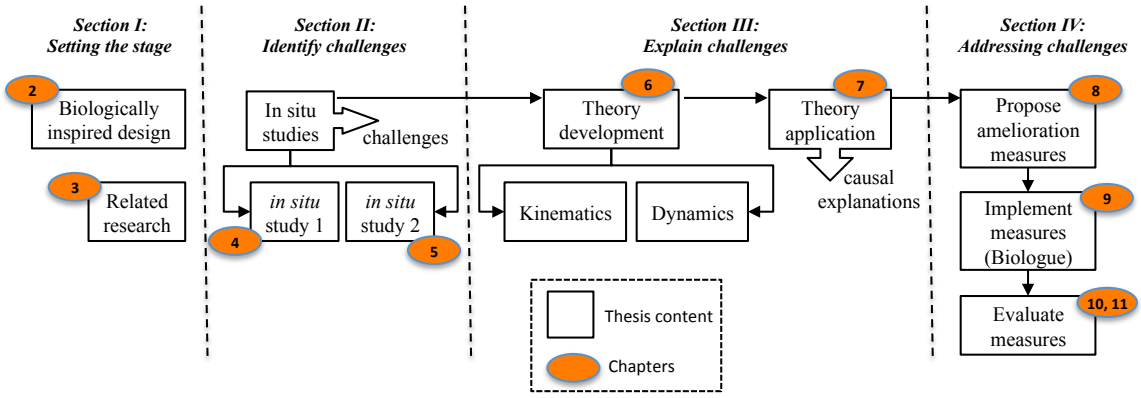
- ***RQ4: How does the presence of an online information environment with features grounded in the hypotheses associated with RQ3 change the dynamics of online bio-inspiration seeking in order to ameliorate the identified challenges?*** *In particular:*
    - ***RQ4.1: To what extent does changing the indexing and access mechanism from keyword-based to functional model-based impact the rate of encountering the relevant information resources?***
    - ***RQ4.2: To what extent does including visual overviews derived from functional models in proximal cues impact the rate of recognition errors?***
    - ***RQ4.3: To what extent does having functional models in addition to textual descriptions of biological systems impact designers' understanding of those systems?***

In order to address RQ4, I conducted a series of evaluation studies, including three experimental studies to evaluate RQ4.1, RQ4.2, and RQ4.3, respectively, and a pilot deployment study in which I deployed Biologue in Fall 2010 in the classroom context of ME/ISyE/MSE/PTFe/BIOL 4803, the same context that was initially used to conduct my *in situ* studies. Details of these studies can be found in Chapters 10 and 11.

This research is still in its early stages, but the trends are encouraging, which seem to suggest that the proposed measures has the potential to change the dynamics of

online bio-inspiration seeking in favor of ameliorating the identified challenges of online bio-inspiration seeking.

Thesis organization



The overall organization of the thesis is shown in the above thesis roadmap. This organization consists of four sections: Setting the stage, Identify challenges, Explain challenges, and Address challenges. Chapter 2 and 3 in "Setting the stage" section discusses the paradigm of biologically inspired design and related research respectively. Chapters 4 and 5 in "Identify challenges" section present the two *in situ* studies through which the challenges of online bio-inspiration seeking are identified. In "Explain challenges" section, Chapter 6 develops a theory of online bio-inspiration seeking and Chapter 7 used that theory to explain the identified challenges. In "Address challenges" section, Chapter 8 hypothesizes about measures that can be taken to ameliorate the challenges, Chapter 9 implements those hypotheses in technology (Biologue), and Chapters 10 and 11 present studies undertaken to evaluate those hypothesized measures.

SECTION 1: SETTING THE STAGE

Section Summary

This section introduces the biologically inspired design paradigm in Chapter 2 and describes my personal motivation for focusing on this paradigm. Chapter 2 also introduces the task of bio-inspiration seeking and makes a case for why this is such an important task in the larger context of biologically inspired design practice. It also makes a case for why this is an intellectually challenging task. Given the manifest importance of this task coupled with the challenging nature of this task, it becomes imperative for biologically inspired design research to gain an in-depth understanding of the nature of this task and propose measures that can be taken to better support designers engaged in this task.

Chapter 3 of this section reviews existing work in the domain of biologically inspired design research. This review indicates that prior theoretical and empirical work in this area has not paid sufficient attention to understanding the nature of the bio-inspiration seeking task. We know very little about the *in situ* practices of designers engaged in this task, nor do we sufficiently understand its underlying processes or 'mechanisms.' Yet, a majority of existing technology-building efforts in the domain of biologically inspired design research has focused on developing tools and techniques for aiding designers engaged in the task. Therefore, there exists a gap between the research of biologically inspired design practice and the technology-building efforts for aiding the practice. Symptomatic of this gap, the current technology-building efforts tend to be technology-centric (as opposed to human-centric), whose design and development are craft-driven (as opposed to theory-driven). In this dissertation, I aim to bridge this gap by adopting a methodology that encourages technology development that is grounded in theoretical understanding of the task, which in turn grounded in *in situ* studies of the task.

CHAPTER 2

BIOLOGICALLY INSPIRED DESIGN

We're going to see in surprisingly short order that biological inspiration and biological processes will be central to engineering real systems. It's going to lead to a new era in engineering.

Charles Vest, "Voices: What's Next in Science," New York Times (2010)

A Brief Introduction to the Paradigm

Consider the following scenario: one night in 1944, Swiss engineer and inventor George de Mestral was frustrated trying to free a stuck zipper on his wife's dress and wondered if he can invent a better fastener. Two weeks later, after walking his dog through the woods, he paused to remove burrs from his pants and from his dog's fur. He noticed how "clever" and efficient burrs were at hooking onto anything they touched. He then examined burrs under the microscope in his workshop and within minutes he sketched what he called "locking tape," and the idea of Velcro was born (Haven, 2006).

A similar story can be told about Dr. Frank E. Fish (Ashley, 2004). One day in the early 1980s he noticed a small statue of a humpback whale in a Boston sculpture gallery. On closer examination, he saw that the creature's large, wing-like pectoral flippers were studded with evenly spaced bumps along their leading edges. As an expert in hydrodynamics, Dr. Fish was taken by surprise. It belied common scientific knowledge that wings and flippers generally needed to have smooth and streamlined front edges. He knew of no cetacean flippers, fish fins, avian wings, or manmade wings for that matter, that bore such odd features - all of those have smooth front edges. He mentioned this to his wife and conjectured aloud that the artist must have made a mistake. The storeowner, overhearing Fish's comments and knowing the sculptor's meticulous attention to detail, soon produced a photograph that clearly showed the humpback's lumpy flippers. But if

the artist was right then at least part of the science of fluid dynamics was wrong. How could that be? This incident became the basis for Dr. Fish's research for the next couple of years. His research found that whereas sheets of water flowing over smooth flippers break up into myriad turbulent vortices as they cross the flipper, sheets of water passing through a humpback's tubercles maintain even channels of fast-moving water, allowing humpbacks to keep their "grip" on the water at sharper angles and turn tighter corners, even at low speeds. Wind tunnel tests of model humpback fins with and without tubercles demonstrated the aerodynamic improvements tubercles make, such as an 8% improvement in lift and 32% reduction in drag, as well as allowing for a 40% increase in angle of attack over smooth flippers before stalling. A company called WhalePower is applying the lessons learned from humpback whales to the design of wind turbines to increase their efficiency, while this technology also has enormous potential to improve the safety and performance of airplanes, fans, and much more.

These two episodes are classic examples of *biologically inspired design* (Vincent & Mann 2002), an approach to designing technology by adapting the knowledge of (the workings of) biological systems. Although bio-inspiration in the two cases highlighted above was incidental, looking to nature for ideation is becoming more deliberate and proactive as the practice of biologically inspired design is gaining more mainstream attention in engineering and scientific community. Proponents of biologically inspired design offer several arguments for why they consider this approach to design beneficial. The two most prominent ones are the *innovation* argument and the *sustainability* argument.

Let us first examine the *innovation* argument. Biological systems occur on the surface of the same planet as human technologies, enduring the same physical and chemical limitations and must use the same building blocks to work with. Nature has already dealt with many of the same problems that many human technologies try to address. But nature copes and invents in fundamentally different ways. As a result,

manmade technologies and biological systems differ dramatically, extensively, and pervasively. In those cases where nature's capabilities are superior to human capabilities, understanding and adapting many of its solution principles can significantly improve our technology (Bar-Cohen, 2005; Vincent, 2001). In other cases, we at least have an alternate "technology" as a mirror in which to view our own.

The potential for innovation through biologically inspired design implicitly relies on the fundamental differences between the nature of biological and technological systems. If not for those differences, the bio-inspired approach would probably lead to technologies that would be less distinguishable from those developed using conventional methods, undermining this approach's potential for producing novel solutions. So, how different are natural solutions compared to human engineered solutions?

At some very basic level all of us recognize how different the products of humans and of nature are. But Steven Vogel's (1998) pioneering work examines this question more methodically. Using several case studies, he first compares the broad structural differences between these two classes of systems and concludes that biological systems differ extensively from technological systems both geometrically and morphologically. He then compares them across broad classes of mechanical functions (e.g., adhesion, pushing, pulling, supporting) and how those functions are accomplished. Here again, he concludes that the means by which those functions are achieved are very different in biological and technological systems. Julian Vincent (Vincent et. al. 2006), on the other hand, takes a quantitative approach to answer the same question. Having analyzed 500 biological systems across 270 functions and comparing them with a similar analysis of engineering systems, he comes to the same conclusion as Vogel. His analysis shows that, for instance, at size levels of up to 1m, where most technology is sited, the most important variable for the solution of a problem is the manipulation of energy usage (up to 60% of the time), closely followed by the use of material. Thus, faced with an engineering problem, our tendency is to achieve a solution by changing the amount or

type of the material or changing (usually increasing) the energy requirement. But in biology the most important variables for the solution of problems at these scales are information and space.

One beneficial side effects of learning from systems that are so different from human engineered systems is that the design outcomes are likely to be very different from existing technologies as well. This can be observed in numerous case studies of biologically inspired design where the new designs are a radical departure from existing technologies. The following represents a small set of radical innovations that can be attributed to the bio-inspired design approach (Vincent et. al. 2006): Velcro inspired by attachment mechanism of burr seeds, one of the toughest materials using glass fiber in a resin mix inspired by wood (the orientation of cellulose in the walls of the wood cells), self-cleaning surfaces inspired by lotus leaf, anti-reflective surfaces inspired by insect eyes, Bionic car design (by Daimler Chrysler) inspired by boxfish, dry adhesive tape inspired by gecko foot hair, highly accurate hearing aid inspired by fly's auditory, drag-reducing surfaces and materials used to make swimwear and hulls of boats inspired by shark skin, wind turbines inspired by humpback whale flippers, etc.

Now let us look at the *sustainability* argument. Nature has not only already solved many of the problems we are grappling with, but has also found out what works, what is appropriate, and most importantly, what lasts here on Earth. After 3.8 billion years of research and development, failures are fossils, and what surrounds us is the secret to sustained survival.

Most human technology takes a "heat, beat, and treat" approach to producing technology. Take Kevlar for instance - it one of the strongest and toughest manmade materials. It requires pouring petroleum-derived chemicals into a pressurized vat of concentrated sulfuric acid, and requires boiling it at several hundred degrees Fahrenheit. It is then subjected to high pressures to force the fibers into alignment as it is drawn out. The energy input is extreme and the toxic byproducts are odious. Nature, on the other

hand, takes a different approach. Because an organism makes similar tough materials like bone or collagen or silk in its own body, the "heat, beat, and treat" approach is not feasible. A spider, for instance, produces a waterproof silk that is superior to Kevlar in toughness and elasticity. Ounce for ounce, spider silk is five times stronger than steel. But the spider produces it in water, at room temperature, using no high heats, chemicals, or pressures. Because organisms have managed to do a lot of things we want to do, without relying on fossil fuels or producing polluting byproducts, what better models could there be to emulate? Although no comprehensive studies examining if biologically inspired design approach naturally leads to more sustainable technology are presented to date, several biological anecdotes presented in favor of the sustainability argument for biologically inspired design are quite compelling.

My encounter with biologically inspired design

At the time that I came across biologically inspired design, I was interested in understanding the cognitive basis of creativity and innovation in the context of design. Because analogy is often linked with creative thought, my goal at that time was to investigate the nature of analogies in creative design.

At the same time, I had become aware of an important methodological shift that had occurred in cognitive science. This shift placed a greater emphasis on observational studies of cognitive phenomena conducted in naturalistic settings. Attempting to account for the role of the environment (social, cultural, material) in shaping and participating in cognitive phenomena was one of the hallmarks of this new "environmental perspective (Nersessian, 2005)" of cognition. In contrast to the conventional view that cognitive processes operated on representations "in the head" alone, the environmental perspectives maintained that cognitive processes could not be isolated and understood separately from the contexts in which they occur.

I had also become aware of the tremendous impact that this shift had had on the domain of scientific cognition. Earlier accounts of scientific cognition tended to focus on topics scoped in limited ways, including computationally tractable reasoning practices gleaned from selective historical cases, studies of expert-novice problem-solving in controlled settings, protocol analyses of scientists or students solving well-defined science problems, etc. By adopting the environmental perspective and by using a new set of methods that paid closer attention to authentic cognitive practices and products of scientists, a new generation of theories of scientific cognition were proposed that were much richer in their explanations and much broader in their scope (e.g., Darden 1991, 1996; Dunbar 1995, 2001; Giere 2004a, 2004b; Gooding, 2004; Nersessian 2005, 2006; Tweney 1989).

This awareness influenced my goal and in turn exposed the need to find a setting that afforded the *in situ* investigation of spontaneous analogizing that occurred "in the wild" during creative design episodes. My advisor and I approached Dr. Wendy Newstetter, a faculty member at the department of biomedical engineering at Georgia Tech. She was known to offer a popular course on biomedical engineering problems that focused on preparing students to tackle complex real-world problems in the area of biomedical engineering. We were hoping to convince her to allow me to use the context of her class to conduct my investigation. Because of the very hands-on nature of this course, there existed a prospect of observing problem solving *in situ* in a design-like domain. But upon learning my objectives, Dr. Newstetter suggested that we should consider a different course for my investigation, namely *ME/ISyE/MS E/PTFe/BIOL 4803: Biologically-Inspired Design*. Apart from an explicit focus on design, this course also had an explicit interdisciplinary focus, bringing biologists and engineers together in order to encourage technological innovation. This course was a relatively recent offering by the Center of Biologically Inspired Design (CBID), an interdisciplinary center for research established with the goal of facilitating biologically inspired design research and

education at Georgia Tech. Dr. Newstetter also mentioned that CBID had recently hosted the 2006 International Symposium for Biologically Inspired Design in Science and Engineering. She handed us the proceedings of that symposium and asked us to get in touch with Prof. Jeannette Yen, who in addition to being the director of CBID was also the primary instructor of the suggested course. We then met Prof. Yen and Prof. Weissburg (co-director of CBID) and thus began my association with biologically inspired design.

In the course of my interactions with Prof. Yen over the next several months, my general awareness of this domain increased. I also became aware of the various biologically inspired design projects taking place at Georgia Tech. I learnt more about the course that she taught every year. Finally, Prof. Yen also introduced me to key biologically inspired design literature. In this period, I uncovered a number of interesting aspects of biologically inspired design that motivated me to choose this domain.

The first aspect of biologically inspired design that I was drawn to was the emphasis placed on innovation. I learnt that this community put innovation front and center in their activities, highlighting the importance placed on non-routine design. This meant that biologically inspired design provided a fertile ground to observe and study human creativity and innovation in the context of design. This strongly aligned with my primary research goals.

The second aspect that I was drawn to was the relative nascency of the field of biologically inspired design. Although people have historically looked to the nature for inspiration for thousands of years, the systematic practice of biologically inspired design was a relatively new phenomenon. This practice was only recently gaining more mainstream attention in both engineering and scientific communities. This meant two kinds of opportunities for design researchers like me: (1) an opportunity to study a novel and emerging design practice (as opposed to more established practices of engineering

design, architecture, etc.), and (2) an opportunity to have a greater impact on the emerging field through our theories, tools, and methodologies.

The third aspect of biologically inspired design that I was drawn to was the foundational role played by cross-domain analogies in this domain. The dominance of cross-domain analogies between the technologies that were the targets of design and the biological systems that inspired their design solutions was evident very early on. I came across article upon article reporting the design of a device or a material based on some analogous biological system. This aspect was very interesting because it seemingly defied many findings from both experimental and *in situ* studies of human analogizing. Experimental studies largely indicated that human subjects have trouble recalling sources from memory that are very dissimilar (yet analogous) to the target (Holyoak & Thagard, 1995; Holyoak & Koh, 1987), implying that deliberate, spontaneous cross-domain analogizing was more of an exception rather than the norm. Similarly, *in vivo* studies of the nature of analogies in scientific cognition also indicated that generative cross-domain analogies were relatively rare compared to within-domain analogies (Dunbar 2001). Therefore, a design community of practice predicated on cross-domain analogizing for its success was a very curious phenomenon.

The fourth interesting aspect of analogies in biologically inspired design pertained to the knowledge status of target and source domains. Analogical reasoning is typically characterized as drawing inferences about a less-understood target domain in terms of a better-understood source domain. But, the converse of this characterization appeared to be true in the case of biologically inspired design: engineers are typically more familiar with the target domain of engineering and less familiar with the source domain of biology. To me, this raised an interesting question of how analogical reasoning can be carried out in situations where the knowledge of source analogues are either sparse or completely missing form the analogist's mind.

To summarize, although I came across this domain through serendipity, my motivation behind pursuing biologically inspired design as the domain of my study was because it (1) was new and exciting, (2) provided a unique window into design creativity and innovation, (3) offered a chance to have significant impact through my research, (4) had interesting characteristics from an analogy research standpoint. Finally, the presence of a BID community in close proximity also offered me the flexibility to adopt a methodology that required direct participant observation.

The Task of Bio-inspiration Seeking

Biologically inspired design is a very complex activity with many different tasks and sub-tasks. The scope of this particular work is much narrow and focuses primarily on one particular task associated with biologically inspired design. Given a target design problem, the task of finding relevant biological systems to emulate is one of the key initial steps in biologically inspired design. This task is often referred to as the task of *seeking bio-inspiration*. In the context of this task, a biological system is considered relevant if the application of the knowledge of its workings can lead to a potentially novel and useful solution to the target design problem.

The task of seeking bio-inspiration is one of the most important tasks in the context of biologically inspired design because the design outcomes are largely determined by which biological system(s) is chosen as an outcome of this task. For instance, consider the design of a bio-inspired seawater desalination technology. If this design is inspired by the salt glands of certain marine animals (e.g., penguins, sea turtles) that help remove excess salt ingested by these animals, then the engineered desalination system would include a single membrane mechanism based on active transport. On the other hand, if this design is inspired by the intestine model of transporting water against concentration gradient, then the engineered system would include a two membrane-based passive mechanism based on a combination of reverse and forward osmosis. Finally, if

the design is inspired by fog-basking beetles or camel noses, then the engineered system would use a thermal approach to desalination that uses coupled evaporators and condensers to produce distilled water. Different sources of inspiration for the same problem can lead to radically different solutions. But not all solutions are created equal. Therefore, the process of finding and choosing the source of inspiration deserves special consideration.

The task of bio-inspiration seeking can also be a very intellectually challenging task. There are an estimated 5 to 15 million species of biological organisms. If one takes into account different levels of organization of biological systems like cellular-, organ-, and ecosystem-levels, then this estimated number of biological systems increases by an order of magnitude or more. Furthermore, novice bio-inspired designers coming from engineering are not familiar with the extent, scope, and richness of biology. They may be aware of only a small fraction of this vast space of biological systems that can be drawn upon in order to develop their design solutions. The near limitless availability of biological systems to draw upon coupled with designers' fractional knowledge of this vast domain of biological systems makes bio-inspiration seeking an intellectually challenging task.

The problematic nature of this task has not gone unnoticed among the biologically inspired design community. In fact, some researchers within this community consider the problem of facilitating the pairing of design problems with useful biological systems as one of the grand challenges of the biologically inspired design agenda (Vincent & Mann, 2002).

If addressing this problem has not been more in the front and center in biologically inspired design research, it is because the transfer of information and concepts from biology to engineering in the context of biologically inspired design has to date been mostly incidental or solution-driven. Incidental here means that the origin of the source of inspiration is either serendipitous or happens through ad hoc associations

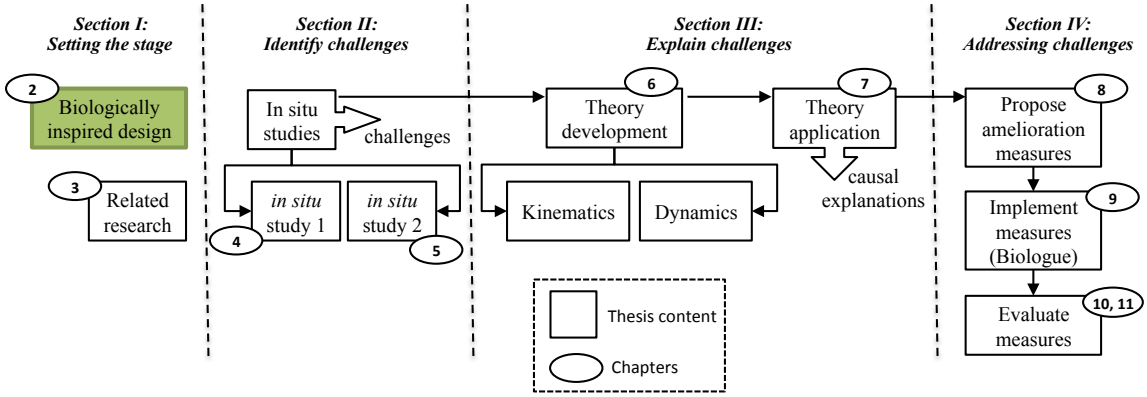
between people. Solution-driven implies that the problem-solving process goes from solutions to problems rather than other way around. In other words, it begins with a biological source and looks for human problems to apply this solution to. For instance, I might be a biologist who has deep knowledge of crab sensory mechanisms. Then I might think about how I can apply this knowledge to develop better sensing technologies.

But things are gradually changing in the arena of biologically inspired design practice. There is now available evidence to show that the biologically inspired design movement is gaining momentum. Bosner and Vincent (2006) cite two metrics that are indicative of the increased attention and adoption this approach is gaining: first, the 'biologically-inspired' patent stock growth and two, the rapid growth in the membership of the BIONIS network during the initial 3-year funding. As more and more designers try to adopt biologically inspired approach to design, the likelihood of more number of people approaching it from a problem-driven side is bound to increase. Which means that more and more people are likely to proactively seek out sources of inspiration rather than staring with a solution or wait for chance encounter with biology. This implies that the problem of problem-driven bio-inspiration seeking will surface more prominently. And for the biologically inspired design paradigm to continue to flourish, the challenges associated with the task of seeking bio-inspiration have to be addressed in a serious manner.

Currently, the problem of facilitating the task of seeking bio-inspiration is an open research problem in the biologically inspired design research community. Researchers are developing various kinds of tools in order to facilitate this task. Biologically inspired designers are being made available information systems of varying degrees of sophistication that provides access to the "right" biological information. However, these tool-building efforts lack a proper theoretical foundation; current tools are a result of pragmatic approaches based on either craft knowledge or empirical research. Principled tool development, evaluation, and comparison are consequently impaired. Furthermore,

these tool-building efforts lack a human-centered focus – they are technology driven, with human issues and concerns being an add-on rather than the primary engine of change. As a result, users have to cope with what the developers have built and are left entangled with tool developer's view of the world.

Chapter Summary



This chapter introduced the biologically inspired design paradigm and described my personal motivation for focusing on this paradigm. It also introduced the task of bio-inspiration seeking and made a case for why this is such an important task in the larger context of biologically inspired design practice. It also made a case for why this is an intellectually challenging task. Given the manifest importance of this task coupled with the challenging nature of this task, it becomes imperative for biologically inspired design research to gain an in-depth understanding of the nature of this task and propose measures that can be taken to better support designers engaged in this task.

CHAPTER 3

RELATED RESEARCH

Instances of Biologically Inspired Design

I begin this review with the kind of research that one typically encounters in the domain of biologically inspired design. Predominantly, literature in this domain is devoted to discussing specific instances of biologically inspired technology or materials that have been developed across a wide range of technical domains. For example, in the field of bio-inspired *optics* one can find examples of computer displays inspired by Morpho butterfly wings (Srinivasarao, 1999, Vukusic & Sambles, 2003); non-reflective surfaces for stealth technology inspired by Hawkmoth wing (Vincent et. al., 2006; Vukusic & Sambles, 2003); advanced photoreceptors and photoconcenrators inspired by Brittlestar (Aizenberg et. al., 2001); improved silica-based optical fibers inspired by Sponge (Sarikaya et. al., 2003); etc. Similarly in the field of bio-inspired locomotion, one can find examples of wall-climbing robots inspired by Gecko locomotion (Arzt et. al., 2003); micro air vehicles inspired by insect (Michelson, 2004) and hummingbird flight (Michelson, 2004); mobile robotics inspired by locomotion of various animals (e.g., Collins et. al., 2005; Ashley, 2003; Full & Koditschek, 1999); underwater oscillatory propulsion systems inspired by fish (Fish, 2006); low-friction surfaces for competitive swimwear inspired by Sharks skin (Dickinson, 1999); etc. Examples in the field of bio-inspired materials include impact-resistant armor materials inspired by Abalone shell (Sarikaya et. al., 2003); fiber construction materials inspired by tree wood, bones (Dabbs & Aksay, 2000) and tendons (Flynn et. al., 2003); structurally resilient glass inspired by Sponge silica (Aizenberg et. al., 2001); porous surfaces for filtration and immunoisolation inspired by diatoms (Baurelein, 2003; Brott et. al., 2001); etc. In the field of bio-sensors, one can find examples of underwater sensing, target acquisition, and

obstacle avoidance technology inspired by Mottled sculpin fish (Gray, 2006); data readout, medical diagnostics, surveillance, and photography technology inspired by Honeybee eye (Jeong et. al., 2006; Lee & Szema, 2005); radioactive plume tracing inspired by moths and lobsters (Farrell et. al., 2005); robotic source tracking algorithms inspired by chemotaxis and C. Elegans (Morse et. al., 1998), etc.

Design Studies of Biologically Inspired Design

While the individual cases of biologically inspired design are interesting in their own right, they are not immediately relevant to the topic of this dissertation. This section reviews literature that can be classified as *design studies of biologically inspired design*, which deals with understanding and furthering the methods, practices, and education of biologically inspired design. This area is relatively new and the body of existing literature is quite limited. The following subsections focus on the theoretical, empirical, and technology-building work related to the topic of bio-inspiration seeking.

Review of theoretical work

Theoretical work in this domain primarily focuses on the *process* accounts of biologically inspired designing. Two kinds of process accounts can be identified. Prescriptive accounts are concerned with prescribing methods for "optimal" biologically inspired designing. They characterize what designers *ought* to do when they engage in biologically inspired design. Descriptive accounts are based on observations of what people *actually* do when they engage in biologically inspired design.

Perhaps one of the most popular processes used to characterize biologically inspired design is the Biomimicry Guild's "Design Spiral" (Biomimicry Institute, 2009). The Design Spiral is a prescriptive account for an iterative design process, where each design iteration informs the next. The basic idea of design spiral has been around in the design literature for some time (e.g., Boehm (1988) in the domain of software design).

Within an iteration, the Design Spiral sub-divides biologically inspired design into six steps: Identify, Interpret, Discover, Abstract, Emulate, and Evaluate. Each step has a set of prescribed actions associated with it. For example, the designer is advised in the Interpret step to "[t]ranslate the design function into functions carried out in nature. Ask 'How does Nature do this function?' 'How does Nature NOT do this function?' Although its origin is unclear, the Design Spiral was presumably derived from best practices in other kinds of design.

BioTRIZ (Vincent et. al., 2006) is another recent prescriptive offering in terms of how biologically inspired design ought to be carried out. It was proposed by Julian Vincent at the University of Bath. It is derived from the earlier theory of engineering invention known as TRIZ (Altshuller, 1984). The TRIZ theory begins with a repository of design cases with known solutions, where each case is indexed by contradictions that arose in the original design situation. Additionally solutions to these design cases are classified into a known set of inventive principles for resolving conflicts. When the designer is presented with a design problem, she reformulates the problem to identify certain key contradictions in the requirements of the design. For each contradiction, she is reminded of a general inventive principle that is applicable for resolving that conflict. In addition to suggesting the essence of a solution for resolving that conflict, the inventive principle also points to a number of cases in which that general principle was instantiated. Vincent et al. (2006) developed a modified version of TRIZ called BioTRIZ specifically for biologically inspired design. The primary difference between the two theories is a change in the features that compose the contradiction matrix to include six "operational fields": substance, structure, space, time, energy, and information. Again, BioTRIZ is a prescriptive account of biologically inspired design, derived from best practices in engineering design.

In our lab, on the other hand, we take a descriptive approach to characterizing biologically inspired design. We have developed a series of increasingly complex information-processing accounts of biologically inspired design, which is partly discussed in this thesis later on. These accounts are based on different kinds of studies of biologically inspired design including direct observational studies, participatory observational studies, specific case studies, as well as analysis of a corpus of projects. Our past studies have yielded the following insights (Helms, Vattam & Goel, 2009). (1) Biologically inspired design engages cross-domain analogies. (2) Problems and solutions in biologically inspired design co-evolve. (3) Problem decomposition is a fundamental process of biologically inspired design. (4) Biologically inspired design often involves compound analogies, entailing a complex interplay between the processes of problem decomposition and the processes of analogical retrieval. (5) Biologically inspired design entails two distinct but related processes: problem-driven analogies and solution-based analogies. We have taken those findings and created an information-processing task model of the process of biologically inspired designing (Vattam, Helms & Goel, 2010). We have also analyzed what makes biologically inspired design a new design paradigm, and, in particular, how the process of biologically inspired design differs from other kinds of design. Design spiral and design matrix accounts view the process as fundamentally the same. In contrast, our task model of biologically inspired design suggests that it differs from other kinds of design in the use of cross-domain analogies, the use of compound analogies, and the use of both problem-driven and solution-based analogies. Parallel to the work presented in this dissertation, we are presently investigating the methodological, technological and pedagogical implications of our task model of biologically inspired design.

Review of empirical work

Empirical work in this domain primarily focuses on the *content* and *representation* of knowledge as they relate to biologically inspired designing. In most cases, research questions are posed assuming that biologically inspired design is an instance of analogical problem solving. The preferred method of empirical investigation tends to be controlled studies conducted in laboratory settings.

One of the earliest empirical investigations of biologically inspired design was reported by Li Shu and her colleagues at the University of Toronto. In one study (Mak & Shu, 2008), engineering students were asked to generate concepts for a simple problem using short text descriptions of biological phenomena as stimuli. They then noted that concepts were generated based on either superficial similarity (where only the biological form is carried over), or deep relational similarity (only the biological mechanisms are carried over), or both. This probably represents the earliest work where a connection between biologically inspired design and a psychological theory of analogy is made. In a second similar study (Vakili et. al., 2007), they also found that there was a tendency for designers to fixate on certain biological mechanisms and force fit those to problems in ways there were inconsistent or impractical.

Chakraborti and his colleagues at the Indian Institute of Science also speculated a connection between biologically inspired design and analogy (Chakrabarti et. al., 2005; Sarkar et. al., 2008). Unfortunately, this connection was not developed any further at that time. Instead, they shifted their focus to questions dealing with knowledge representation. They hypothesized that both the nature of content and representation of biological sources influence the extent to which they trigger inspiration. They studied the effect of representation of triggers on ideas generated by six design engineers while trying to solve a given problem. A variety of representations (video/animation and audio, text, explanation, and others) that are potentially useful to designers for five pre-specified triggers were administered to each designer, who generated ideas in response to each trigger–representation combination individually. The effect of representations of these

triggers on the content and representation of the solutions generated by the design engineers was studied. Their results showed significant influence of the representation of the triggers on the representations, number, and quality of the resulting ideas that were generated.

Following their work on representation, Chakraborti and his colleagues turned their attention to the nature of analogical transfer in biologically inspired design (Sartori, Pal & Chakrabarti, 2010). Their empirical studies of designers from two different countries carrying out biologically inspired design shows that transfer typically takes place at four levels of abstraction: state change, organ, attributes, and parts. They also found that, when unaided, biologically inspired designing was dominated by transfer at part, attributes and organ levels, while little transfer takes place at state change level. While this provides greater reliability for design outcomes, it reduces the novelty of the designs that can be produced. Based on this finding they develop a new set of guidelines to support the analogical transfer process. These guidelines are then used to inform a systematic support system called "Idea-Inspire" that provides analogically relevant stimuli. A comparative study also showed a steady and significant increase in the total number transferred designs when aided by their guidelines, as well as a shift in level of abstraction, favoring state change and organ levels.

Research by Linsey et al. explores the nature and role of representations in the context of design-by-analogy (Linsey, Wood & Markman, 2008; Linsey, Markman and Wood, 2008). Although their claims are not about biologically inspired design *per se*, their empirical work uses a significant number of cases from biologically inspired design, implying that they view biologically inspired design as a kind of design-by-analogy and that their claims are applicable to biologically inspired design. Their claim is that representation clearly matters and seeking improved representations will enhance the analogy-making process. A more general semantic description of a product allows for a greater chance of using a previously experienced product or (a biological system) as a

source analogue later. The results and insights gained from their design experiments support the assertion that the form of concept representation is important in the cognitive analogy formation process. Similarly, the form of functional representation used for analogy searching in natural language-based databases and the metric for evaluating the analogy between concepts is critical to the success of the systematic analogy search methodology. The semantic functional representation should enable easier access to information stored in repositories such as patent archives and websites (Linsey et. al., 2006).

Review of technology-building work

A majority of the technology-building work in the area of biologically inspired design has focused on aiding the task of bio-inspiration seeking. The problem of facilitating the task of seeking bio-inspiration is acknowledged as an important open problem by the biologically inspired design research community (Vincent et. al., 2006; Bar-Cohen, 2006).

There are predominantly two approaches for facilitating the task of seeking bio-inspiration. One straightforward approach is to bring biologists to the design table by forming multidisciplinary design teams of engineers and biologists. In doing so, it is hoped that biologists will serve as beacons and point engineers in the direction of biological systems that may be relevant to the design challenge at hand. This approach, albeit based on a naïve view of multidisciplinary team dynamics, is considered a standard practice for getting help with the task of seeking bio-inspiration. This approach, which relies on human-human interaction, is limited by the availability and the cost of biological expertise.

Another approach for facilitating the task of seeking bio-inspiration is through human-information interaction. In this approach, which relies heavily on the use of information technology, designers are made available information tools of varying

degrees of sophistication that provides access to the "right" biological information. I refer to these information technologies as *bio-inspiration seeking support technologies*. A large part of this chapter is dedicated to reviewing and analyzing current research efforts related to bio-inspiration seeking support technologies. In this chapter three mains areas of research related to such technologies are discussed.

These tools include organized databases, information retrieval systems, engineering-to-biology thesaurus, and different varieties of knowledge-based systems. The details of these bio-inspiration seeking tools are discussed below.

Organized Databases

One of the earliest tools developed to support the task of seeking bio-inspiration came out from the Biomimicry Institute (2009). It was known as the Biomimicry Portal when it started (http://database.portal.modwest.com/). This online portal was intended to support the task of seeking bio-inspiration. This tool emerged from within the community based on the recognition that practitioners of biologically inspired design express one consistent need - access to relevant biological information. It was intended to be a place where designers, architects, and engineers, etc. could search biological information to find ideas that potentially solve their design challenges. This portal contained biological information organized based on the following features. (1) *Challenges* are human design problems that need solutions. (2) *Strategies* are potential biological solutions to those problems. (3) *Organisms* describe specific organisms, listing their taxonomic categorization, a description of what the organism has/does that might be inspiring, and data on the organism's environment. (4) *Products* are descriptions of biomimetic products, including company names and contact information and product availability. A basic search tool allowed users to search this information using keyword. An advanced search tool allowed users to narrow down search to the particular information-types mentioned above. A browse tool allowed users to have a hierarchical table-of-contents

view of the information. Due to lack of published evaluation information, it is difficult to know the effectiveness of this system. The Biomimicry Portal was later replaced by a more modern AskNature (2012) application. While retaining the database of biological information, AskNature incorporated a social networking infrastructure where users could signup with AskNature and participate in an online community formed around the education and practice of biologically inspired designers. In addition to the goal of connecting designers with the right biological information, AskNature also intended to connect designers with other designers and/or biologists within this online community who shared common or complimentary interests.

Julian Vincent is also implementing a tool to support his BioTRIZ (Vincent et. al., 2006) approach using programming techniques developed for the Semantic Web, based on the Resource Description Framework (RDF), a formal language for describing structured information (http://wiki.bath.ac.uk/display/OOB). RDF allows exchange of information on the Web and so lends itself to a communal resource. A recent development of RDF is OWL2, a Web Ontology Language implemented in a Public Domain program, Protégé-OWL, which provides the database infrastructure for his database. BioTRIZ database is intended to organize and analyze information from biological publications. Using Protégé he has generated an ontology that describes the logical connections between the types of data used and populated it with biological examples culled from the literature, and analyzed to show where they converge with each other and with human-made technology.

Information Retrieval Systems

Li Shu and her colleagues at the University of Toronto have also tried to address the problem of facilitating the task of seeking bio-inspiration (Mak & Shu, 2008). In contrast to the database approach, their method adopts a natural language processing approach, hoping to take advantage of the abundant biological information that is already

available in books, journals, and so forth, by performing keyword searches on these existing natural-language sources. One challenge of retrieving relevant cross-domain information using keyword search involves differences in domain vocabularies or lexicons. The problems of differing lexicons are addressed through exploring and analyzing the corpus itself. They provide a statistical "bridging" approach to taking a keyword (specifically, a verb) that the designer might use to search the corpus (e.g., clean) and use that to find biologically connotative verbs (e.g., defend, capture) which is likely to yield more relevant information from the biology corpus.

The problem of differences in domain vocabularies between engineering and biology is tackled slightly differently by Nagel and others (Nagel, Stone & McAdams, 2010). They address this problem by providing an engineering-to-biology thesaurus, which lists biological correspondent terms that an engineer, who is using a function-based design approach, might encounter. Biological terms in the thesaurus are correlated to the engineering domain through pairing with a synonymous function or flow term of the Functional Basis lexicon (Nagel et. al., 2008), which suggests that this thesaurus is intended to be used in conjunction with Function Basis modeling. After creating an initial model of the system being designed, the terms from this model can be used with the thesaurus to find biological correspondents. The population of the engineering-to-biology thesaurus was achieved through functional word searches of a biological textbook that covered a broad range of topics, described as an organized verb-noun search. Chosen words were determined by their macro-relevancy, which is identified by frequency of use. Functional Basis functions (verbs) were utilized for searching the biological textbook to extract biologically connotative words (nouns) that an engineering designer interested in function based design might encounter.

Knowledge-based Systems

One of the recurring themes in the engineering design research community is the utilization of systematic methods or tools that aid in the engineering design process. These are used for supporting a number of design activities including concept generation in the early stages of design. A small number of researchers from this community have turned their attention to biologically inspired design and have proposed design methodologies and supporting tools. Most of their products are based on model-based design techniques to facilitate biologically inspired design. Systems modeling using a standard formalism is recognized as a way to connect nature and engineering through a commonality. Using functional representation and abstraction to describe biological systems presents the natural systems in an engineering context and allows designers to make connections between biological and engineered systems. Thus, the biological information is accessible to engineering designers with varying biological knowledge, but a common understanding of engineering design methodologies. By creating a bridge between the two domains through the perspective of function, it is expected that engineers can better leverage the biological information. The following are some of the main efforts undertaken along this direction.

Chakrabarti and his colleagues take a knowledge-based approach to assisting designers in the task of seeking bio-inspiration (Chakrabarti et. al., 2005). They have developed a knowledge-based system called Idea-Inspire, which stores information about both natural and manmade systems using a common language for representing these systems and their functionality. It provides procedures for interactive retrieval of potential sources of inspiration and generation of alternative ideas for solving a given design problem. The common representation language is referred to as the SAPPhIRE model of causality. It allows the function, behavior, and structure of a system to be linked to each other in a way common for both natural and artificial systems, and allows describing these at various levels of abstraction. Both function and behavior of a system are taken to be descriptions of what a system does, except that function is intentional and

at a higher level of abstraction than its behavior, which can be taken as the way in which the function is achieved. Structure is described by the elements and interfaces of which the system and its immediate, interacting environment are made. At the core of behavior of a system are changes of the state of the system, and how these are brought about by the right contexts formed by the properties of the system and its environment, and inputs from these, in order to activate the physical effects necessary to effect the change of state. The usage of Idea-Inspire is envisioned as follows. If the design problem is well defined (problem can be precisely captured using the constructs of the representation), then the designer can provide this design problem as an input to the system and use the reasoning procedures of the software for automated search for solutions. On the other hand, if the problem is not well defined, the designer can browse the knowledgebase and view information related to biological or artificial mechanisms, get interested in some of these mechanisms, and then use those ideas to solve the problem. Browsing may also help in understanding a problem better, as a designer will be exposed to a wider variety of related yet concrete solutions.

In our earlier work, we have also taken a similar knowledge-based approach with our system called DANE (Design by Analogy to Nature Engine) (Vattam et. al., 2010; Wiltgen, Vattam & Goel, 2011). But our approach goes beyond pointing designers to relevant biological systems and also includes the goal of supporting engineers in understanding those biological systems. DANE contains functional models of biological systems in Structure-Behavior-Function (SBF) formalism. However, the SBF modeling framework, originally created for automated reasoning in AI systems, has been redesigned to facilitate easy, visual authorship and human readability. SBF models of biological systems in DANE are intended to be interactive external representations that facilitate understanding the workings of those systems. Systems in DANE are indexed by system-function pairs, allowing them to be accessed by name (e.g., “lotus leaf cleans self”), by subject (e.g., “lotus leaf”), and/or by verb (e.g., “clean”). Upon selecting a

system-function pair, users are presented with a multi-modal representation of the paired system-function (e.g. the Lotus Leaf SBF model). A system can be represented in text descriptions and images, as well as through visualizations of behavior and structure models. Behavior and structure models are themselves represented as directed graphs, which may be annotated with text descriptions and images. The nodes and edges represent either structural elements and connections (for structure models) or states and transitions (for behavior models), respectively. Additionally, each system is visually connected to other systems with which it shares a sub or super-function relationship. This functional hierarchy is represented as an interactive graph with nodes representing systems and edges representing sub/super relationship.

McAdams, Stone and others have explored a different design formalism to support the task of seeking bio-inspiration (Tinsley et. al., 2007). Their work relies on using the *Functional Basis* (Stone & Wood 2000), a design language consisting of a set of functions and a set of flows that are used to form sub-functions, to model biological systems. Through a number of case studies they demonstrate the feasibility of modeling biological systems using Functional Basis (Stone & Wood, 2000). They also show that once represented in this form, morphological matrices can be used to establish relationships at the functional level between embodied engineering solutions and naturally occurring biological systems. Encouraged by their results they propose to build design repositories containing Function Basis models of biological systems and supporting tools for accessing those models.

Other research by Singh et al. explores transformation principles in biomimetic design (Singh et. al., 2006). A product that can transform to fulfill multiple functions can increase efficiency, reduce cost, and increase weight saving. In Singh's work, a methodology is developed for creating innovative products with broader functionality through the exploration of transformation design principles. The paper details case studies in nature, patents, and products. The three transformation principles deduced from

the case studies are “Expand/Collapse,” “Expose/Cover,” and “Fuse/Divide.” Accompanying these are transformation facilitators, which include “Common Core Structure,” “Composite,” “Conform with Structural Interfaces,” “Flip,” “Function Sharing,” “Furcation,” “Generic Connections,” “Modularity,” “Nesting,” “Shared Power Transmission,” and “Shelling.”

Discussion

Consider the numerous ongoing tool-building efforts related to biologically inspired design. Notice that although most of these efforts are intended to address the problem of facilitating the task of seeking bio-inspiration, they are not grounded in the actual practices of biologically inspired design. The fundamental question “how do designers situated in one domain (engineering) currently find relevant systems form a vast space of available systems that belong to a completely different and mostly unfamiliar domain (biology)” has not been researched in its own right. Therefore, there currently exists a gap between the research of biologically inspired design practice and the technology-building efforts for aiding the practice. Symptomatic of this gap, the current technology-building efforts tend to be technology-centric (as opposed to human-centric), whose design and development are craft-driven (as opposed to theory-driven).

There are two shortcomings associated with the existing approaches to tool building. First, they are theory-thin approaches, lacking sufficient grounding in theories that help analytically and critically comprehend the task they are intended to support. The breadth of theoretical work in biologically inspired design research (as limited as they may be) primarily focuses on the process of design as a whole and not on the task of seeking bio-inspiration *per se*. The level of abstraction of these theoretical accounts is so high that the task of seeking bio-inspiration usually gets abstracted away. If our goal is to develop tools that aid the task of bio-inspiration seeking, it behooves us to base them on

frameworks that bring the task of bio-inspiration seeking more firmly into their theoretical fold.

Second, current approaches belong to a tradition that we might dub as technology-centered design (contrasted with human-centered design), where the development is primarily focused on the technology, with human issues and concerns being an add-on rather than the primary engine of change. Technology-centered design approach at best assumes and at worst ignores the *actual* needs and preferences of the users. This approach prioritizes the attributes of the technology itself and often results in design solutions that are in search of problems. Its limitations have given rise to human-centered design, which recognizes that system design will benefit from the explicit study of the socio-cultural context in which users work. This explicit focus on "context" is missing in current technology-building efforts. For instance, none of the current efforts have ventured to observe and critically study the actual bio-inspiration seeking practices of designers *in situ*. As a result, "what is" has failed to inform "what ought to be" in the current approaches.

Theory-thin Approaches

How and why are tools useful? How can their usefulness be explained? How can they be made more useful? Usefulness is and should be one of the central concerns for tool developers. However, the issue of usefulness becomes problematic particularly when they are meant to work with humans. The reason for this is that their purpose becomes defined, at least in part, in terms of what they do for the user. In the context of the task of seeking bio-inspiration, this is frequently related to human cognition because the task of seeking bio-inspiration is a complex, creative, intellectual activity ('cognition' here is taken in its broader sense to encompass embodied and distributed cognition).

One crucial aim of bio-inspiration seeking tools is to assist and improve the cognitive processes underlying that task. Simply put, a bio-inspiration seeking tool is

considered “good” if they support cognition in the context of that task. There may be many other reasons for why a tool is considered good (computational efficiency, usability, learnability, etc.), but its ability to support cognition is surely a central one. Ultimately, then, the explanation offered for the design of a bio-inspiration seeking tool will need to rest on some account of which of its features assist what cognitive processes and how.

It is important to be able to clearly articulate explanations for why a tool is believed to support bio-inspiration seeker cognition. If the “claims” about a tool are not made explicit, it is extremely difficult to test them, to compare tools, or to reuse design knowledge. Clearly, any claim about the *cognitive support* provided by a tool will be at least partly psychological in nature. So it seems prudent to desire that our rationalizations be firmly grounded in well-received theories from cognitive science.

Unfortunately, current bio-inspiration seeking tools are too rarely analyzed for the cognitive rationales underlying their design. This makes it considerably less clear what generalizable lessons can be drawn from the tool. For some of the key lessons will relate to the cognitive benefits of the tools. To grasp these, we must have “deep” cognitive descriptions of these benefits - not merely “shallow” explanations at the technological level of search algorithms and interface features. The deep cognitive explanations make it possible to generalize the lesson beyond the specific implementation context.

What exactly are some of the practical difficulties associated with theory-thin approaches to interactive tool development? HCI research gives some insight on this issue. The troubles stemming from a lack of a theory-guided research stream for tool development in HCI is discussed at length by Walenstein (2002). A summary of this discussion is summarized in Table 3.1.

Table 3.1: Limitations of theory-thin approaches

Activities	Difficulties with theory-thin approaches
Evaluation and testing	Tool support claims are poorly articulated and tested
	"Whole tool" testing is needed but is burdensome, problematic
	Tool developers are forced to engage in naïve cognitive science
Analysis	Informal analysis suffers from concept- and lexicon- poverty
	Analysis is shallow because deep analysis requires knowledge of psychology or cognitive science
Design	Based on craft knowledge or on trial and error
	Design is affected by theory too late, or not at all

On the other hand, what advantages do (an appropriately chosen) theoretical frameworks bring to the tool-design table? According to Halverson (2002), theoretical frameworks provide *descriptive* power - helps us make sense of and describe the world. This includes describing a work setting as well as critiquing an implementation of technology in that setting. Second, they provide *rhetorical* power - helps us talk about the world by naming important aspects of the conceptual structure and how it maps to the real world. This is both how we describe things to ourselves and how we communicate about it to others. Further, it helps us persuade others to accept our views. Third, they provide *inferential* power - helps us make useful inferences. In some cases those inferences may be about phenomena that we have not yet understood sufficiently to know where or how to look. We may hope that inferences will lead to insights for design. Or we may want to predict the consequences of introducing change into a particular setting. Fourth, they provide *application* power - applying the theory to the real world for essentially pragmatic reasons. Mostly this translates to our need to inform and guide system design. We need to describe and understand the world at the right level of analysis in order to bridge the gap from description to design.

Technology centric design

Most of the current tool-building efforts to aid the task of seeking bio-inspiration are technology centric, with human issues and concerns being an add-on rather than the primary engines of change. For the most part, tools are designed and implemented without fully taking into account designers' cognitive abilities, the ways they perceive and handle information, go about their work and life, create and maintain their social relations, or use their cultural context. That is, tool research and developers often develop computing technologies in relative isolation. This is not a problem in and of itself. It becomes problematic when those tools are interactive tools, intended to work with humans in order to mediate some task, rather than to automate those tasks.

In technology-centered design, system developers specify the requirements for machines, then implement or prototype the requirements, and finally produce devices and software. They then go away, leaving end users to cope with what they have built. Indeed, experience has shown that devices that are designed according to design-then-train philosophy "force users to adapt to the system. The user is entangled with the system that represent the designer's view of the world" (Hoffman, 2002). Many lessons learned over recent decades have pointed toward a need for an alternative approach. These lessons span a range, including insights from well-intended tools that cause user frustration and go unused due to the differences between designers' and users' worldviews. But the lessons also come much closer to home. We all have experienced, for example, the frustrations of learning to use software, advertised and lauded for its new capabilities by those who designed it and are therefore familiar with it. The new capabilities, however, usually require significant relearning, backpedaling, kludging, and workarounds. Bells and whistles often go unused or even unnoticed.

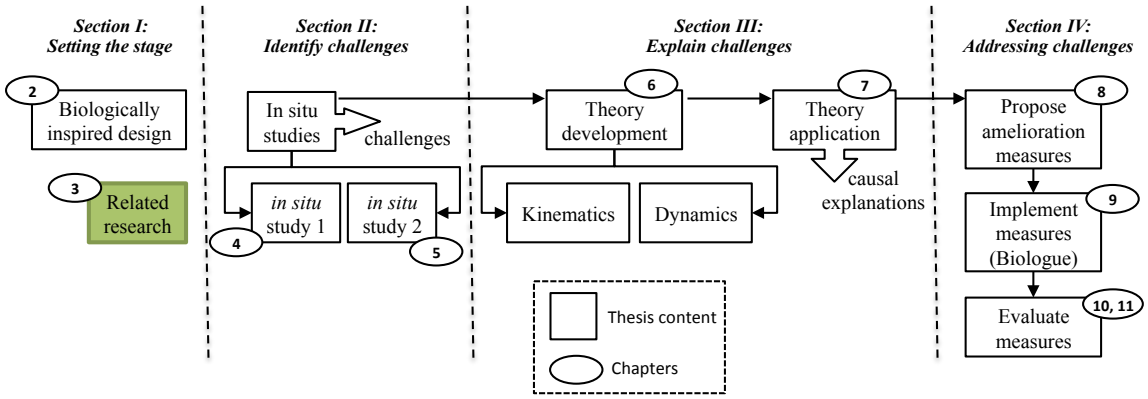
The alternate vision is *human-centered computing* (Kling & Star, 1998; Dertouzos, 2001; Flanagan et. al., 1997; Hoffman, 2004; Hoffman et. al., 2002, 2001), whose hallmark is to focus on domain practitioners, and their field of practice. Human-

centered design depends on a deep analysis of a field's cognitive and collaborative demands and how people work individually and in groups to meet those demands. "The objective is to leverage what we know about human cognitive and collaborative processes to create systems that optimize the *affordances* (direct perception of meanings) and *effectivities* (knowledge-driven actions) for humans" (Scott et. al., 2005, pp. 73).

Another aspect of human-centered computing is a focus against what might be called *Laboratory-Based Design* (Hoffman et. al., 2002). The basic idea is that the tool-building process cannot be conducted by cloistered developers and programmers feeding designs to the user. Rather, tool developers must become field researchers and immerse themselves in the application domain to fully understand domain practice and the context of the prospective designs' use. Contrary to this approach, majority of the afore-reviewed tool-building efforts have not been informed by how the task of seeking bio-inspiration is actually performed in the field.

Although it has been recognized that system design will benefit from explicit study of the context in which users of the designed system work (Nardi, 1992; Halverson, 2002), current efforts neglect to study how designers seek bio-inspiration in naturalistic settings and base their design on a model of user as an unaided individual divorced from a social setting and from supporting artifacts.

Chapter Summary



This chapter reviewed existing work in the domain of biologically inspired design research. This review indicated that prior theoretical and empirical work in this area has not paid sufficient attention to understanding the nature of the bio-inspiration seeking task. We know very little about the *in situ* practices of designers engaged in this task, nor do we sufficiently understand its underlying processes or 'mechanisms.' Yet, a majority of existing technology-building efforts in the domain of biologically inspired design research has focused on developing tools and techniques for aiding designers engaged in the task of bio-inspiration seeking. Therefore, there exists a gap between the research of biologically inspired design practice and the technology-building efforts for aiding the practice. Symptomatic of this gap, the current technology-building efforts tend to be technology-centric (as opposed to human-centric), whose design and development are craft-driven (as opposed to theory-driven). In this dissertation, I aim to bridge this gap by adopting a methodology that encourages technology development that is grounded in theoretical understanding of the task, which in turn grounded in *in situ* studies of the task.

SECTION II: IDENTIFYING THE CHALLENGES

Section Summary

The information presented in this section tries to address the research question:

- ***RQ1: What are the fundamental challenges of online bio-inspiration seeking?***

In order to address RQ1, I conducted two *in situ* studies of biologically inspired design. Both studies were conducted in the context of ME/ISyE/MSE/PTFe/BIOL 4803, a project-based introductory course on biologically inspired design offered in the Fall semester of every year at Georgia Tech. These two studies are reported in Chapters 4 and 5 in this section. The objective here was to gain an understanding of the characteristics and challenges of the online bio-inspiration seeking activity based on studying it in an *in situ* environment.

In the context that I studied, several key aspects characterized that online bio-inspiration activity. First, it involved a search for one or more *cross-domain analogies* between the target technology that was the subject of design and source biological systems, mediated by several kinds of online information environments (predominantly those which gave access to scholarly biology articles like Web of Science, Google Scholar, etc.). Second, it was characterized by the application of unique *strategies* such as "biologizing the problem" and the use of *abstractions* such as functions, mechanisms, principles, constraints, etc., in order to bridge the engineering-biology divide during the search process. Third, it was characterized by a process that was not only collaborative, but consisted of three stages: pre-search stage consisted of team-level activities used to come to a shared understanding of the problem, establish information needs, negotiate division of labor, etc.; during-search stage consisted of individual information-seeking activity in order fulfill the identified information needs; and after-search stage consisted of representation-construction activities and information organization and sharing activities. Fourth, the individual information-seeking process was highly exploratory and

open-ended, took up a lot of designers' time and yielded relatively small number of information resources that contained actually relevant (analogous) biological systems.

In the context that I studied, I also identified some of the challenges of online bio-inspiration seeking.

Thesis 1: Designers engaged in online bio-inspiration seeking face at least three fundamental challenges: First, designers experience a *low rate of encountering relevant information resources* in online environments that they normally rely on. Second, designers experience a *high rate of recognition errors*: they fail to recognize the true relevancy of the information resources that they encounter in those information environments. Third, designers experience *significant difficulty in comprehending information resources* that they recognized as being relevant and struggle to develop conceptual understanding of biological systems discussed therein.

These three issues contribute significantly to lessen the efficiency of online bio-inspiration seeking process, which in turn contributes to the problematic nature of this task for designers. Therefore, it becomes important to understand the causes of these challenges and propose appropriate measures to ameliorate these challenges by targeting those causes.

CHAPTER 5

AN INITIAL STUDY OF BIOLOGICALLY INSPIRED DESIGN

This chapter presents our initial study of biologically inspired design. This study is one of the first *in situ* studies of biologically inspired design. The objective of this study was to gain a better understanding of the processes of biologically inspired design as a design activity, including the process of bio-inspiration seeking. This study was conducted in the context of an undergraduate interdisciplinary course on biologically inspired design at Georgia Tech (ME/ISyE/MSE/PTFe/BIOL 4803). Although this study was conducted in the context of a classroom, the goals of this study were both to understand the nature of biologically inspired design and to identify opportunities for enabling more effective practice of biologically inspired design at large.

The Context of the Study

ME/ISyE/MSE/PTFe/BIOL 4803 is a project-based undergraduate class, in which 45 students, 41 of whom were seniors, work in small teams of 4-5 designers on assigned projects. The class was very interdisciplinary, composed of 6 biologists, 25 biomedical engineers, 7 mechanical engineers, 3 industrial engineers, and 4 from other majors. The projects involve identification of a design problem of interest to the team and conceptualization of a biologically inspired solution to the identified problem. Each team writes a 15-20 page report and makes an oral presentation near the end of the semester. In Fall 2006, ME/ ISyE/MSE/PTFe/BIOL 4803 was jointly taught by six faculty members from Georgia Tech's Schools of Biology, Chemistry, Mechanical Engineering, Industrial & Systems Engineering, and Polymer, Textile and Fiber Engineering. The course also included guest lectures by several prominent researchers from other schools.

The ME/ISyE/MSE/PTFe/BIOL 4803 class was structured into lectures, found object exercises, journal entries, and a final design project. Most lectures focused on exposing the designers to existing biologically inspired design case studies. Other lectures were devoted to the design processes involved in biologically inspired design work: reframing engineering problems in biological terms, functional analysis of a problem, optimization, and the use of analogy in design. Some lectures posed problems for the students to solve in small group exercises.

Although this study was conducted in the context of a classroom setting, we approached the study from a design cognition perspective as opposed to a learning sciences perspective. That is, we were less concerned about the pedagogical approach and the learning outcomes of the course. Although we believe that our research will have implications on the approach and conduct of the course, we were not directly involved in the decision-making regarding the design of the course. From our perspective the classroom provided a setting where we could observe designers engaged in biologically inspired design.

Most instructors and lecturers had many years of practical biologically inspired design experience and focused classroom lectures on sharing their biologically inspired design experience through specific case studies. Most students, although new to biologically inspired design, had previous design experience. Out of the 45 students, at least 32 had taken a course in design and/or participated in design projects as part of their undergraduate education. Throughout this paper, we will refer to the students in the class as designers.

In addition to lectures, classroom activities included regular found object exercises that required designers to bring in biological samples and analyze the solutions employed by these samples. These exercises were intended to expand awareness of biology, provide hands on experience with biological systems, and encourage the designers to dig progressively deeper into the functions of biological systems.

Additionally, journal entries required designers to write about their classroom experiences, including found object discussions, and to document their own design thinking.

The final design project grouped an interdisciplinary team of 4-5 designers together based on interest in similar problems or solutions. Each team had at least one designer with a biology background and a few from different engineering disciplines. Each team identified a problem that could be addressed by a biologically inspired solution, explored a number of solution alternatives, and developed a final solution design based on one or more bio- logically inspired designs. The teams presented their final designs during the final two weeks of class and submitted a final paper, which combined represented a majority of their semester grade.

As observers, we attended all the classroom sessions, collected all course materials, documented lecture content, and observed teacher-designer and designer-designer interactions in the classroom. We also observed a few of the interdisciplinary teams of designers engaged in their design projects. We minimized our intervention, only occasionally asking clarifying questions. Our observations focused on the processes and the products of the designers. In terms of the practices, we observed and documented frequently occurring problem-solving and representational activities of designers. In terms of the design products, we observed and documented the 'design trajectory' - the evolution of the conceptual design over time.

Case Report 1: Project BriteView

The goal of the BriteView project was to design a display screen that was resistant to drowned illumination in bright sunlight and one that is power efficient. The problem was reframed, or "biologized," as: "How do organisms in nature generate bright, crisp colors even in the presence of bright sunlight?" From the reframed problem, designers found three biological sources of inspiration, Morpho butterfly wings, hummingbird (and

duck) feathers, and peacock feathers. Based on the optical properties of each, an initial bio-inspired solution was created based on the Morpho butterfly wings. This solution suggested creating a Christmas tree-like thin-film structure for each pixel that produced structural coloration through the interference effect (the butterfly wings are lined with such Christmas tree- like nano structures). Upon evaluation, designers felt that this solution was infeasible due to the complexity in manufacturing such intricate structures.

Designers chose the humming bird feathers as their next source of inspiration. Although the structural coloration produced by the humming bird feathers is based on the same optical principle as that of the butterfly wings, the hummingbird feathers contain a series of alternating layers of thin-films with different thickness instead of the intricate Christmas tree-like structure. Since simple layering of thin-films is more feasible to implement, this source was selected. At the same time this solution was being developed, designers also considered the structure of peacock feathers (the third source of inspiration). Any solution based on peacock feathers was quickly rejected because they had to contain multi-dimensional structure (as opposed to single-dimensional structure in both butterfly wings and humming bird feathers), which was considered even harder to implement.

Based on the humming bird feathers, the initial solution suggested that each pixel contain a two-layered thin-film structure, each layer having a different thickness. When they initially evaluated this solution, they realized that this solution did not give them the control to dynamically vary the color produced by the pixel, which was crucial for the design of the display. Then they revisited their earlier source of inspiration, the butterfly wing, because they knew that the color that the wing produced was determined by the

length of the air gap between the layers in the Christmas tree-like structures. Varying the length of this air gap would vary the output color. Using this principle they modified their initial solution to include a gap between the two layers filled with air. Now they could move the bottom layer up and down mechanically changing the length of the air gap between the two layers, which in turn effected the color change in the pixel.

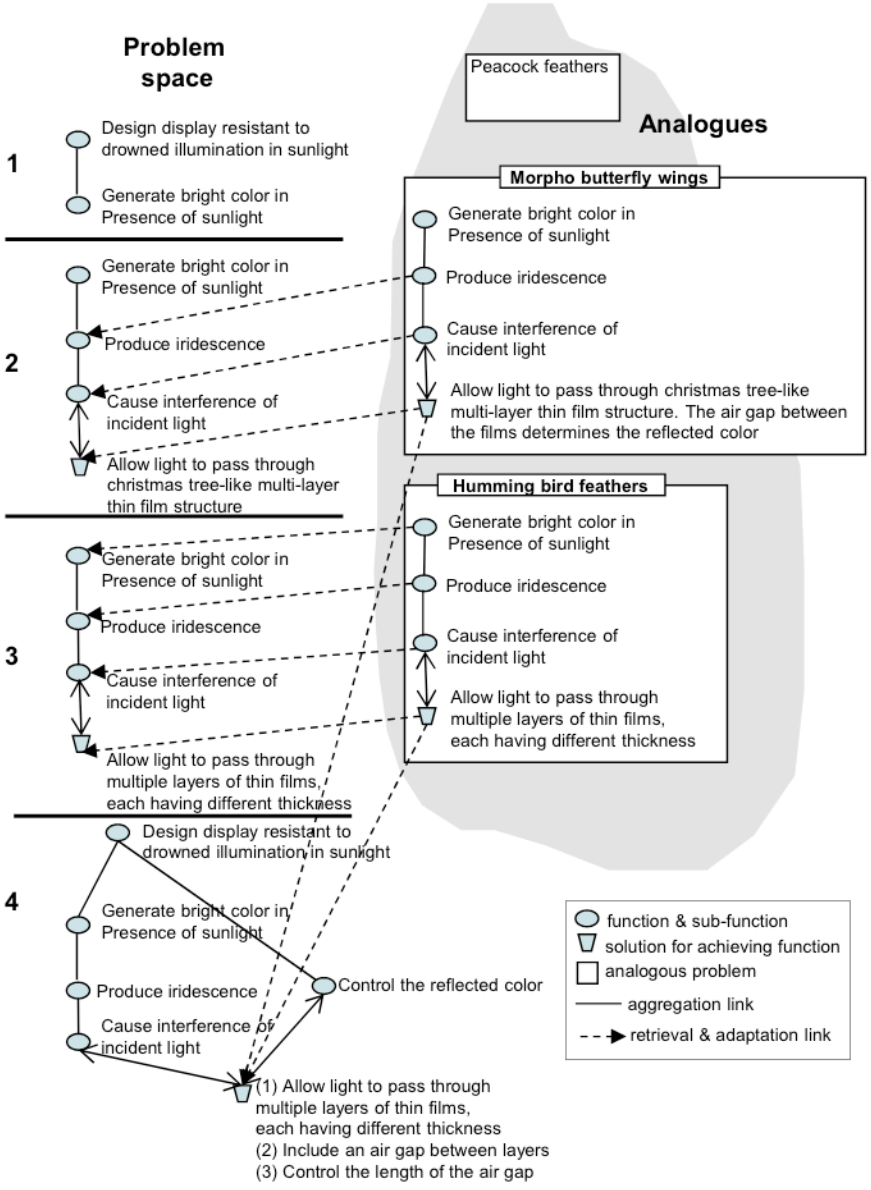


Figure 4:1 Design trajectory of Project BriteView

The Figure 4.1 shows the design evolution of this solution. Step 1 depicts the problem space early in the design. The overall function “design a display” has been decomposed based on the background knowledge and one of the sub-functions “generate bright color” has become the focus. Step 2 shows the initial solution generated based on the first source analogue, Morpho butterfly wings. This solution was evaluated and rejected. In Step 3 another trial design is generated based on the second source analogue, humming bird feathers. This is evaluated and a new function “control the reflected color” is added to the problem space. Step 4 shows the addition of this new function and an improved solution that integrated the idea of air gap (inspired by the Morpho butterfly wing design) into the trial design generated in Step 3.

Case Report 2: Project InvisiBoard

The goal of this project was to conceptualize a new kind of surfboard that prevented the formation of the surfboard and surfer silhouette (which resemble the silhouette of a shark prey when seen from below) to prevent “hit-and-run” shark attacks due to mistaken identity. This problem was biologized as: “how do organisms camouflage themselves in water to pre- vent detection by their predators?” The following biological systems were considered as potential sources of inspiration. (i) Indonesian mimic octopuses are expert camouflage artists. They can mimic various animals based on which predator is close by. Upon studying closely, this source was rejected because the surfboard is a rigid body and does not afford the same flexibility as the body of an octopus. (ii) Bullethead parrotfish uses the principle of pointillism to camouflage themselves. When viewed at close range, the fish appear bright and colorful but when viewed from a further distance, the combination of the complementary colors creates the illusion that the fish is grey-blue. This trick blends the parrotfish into the backlight of the reef, and in essence it disappears. (iii) Pony fish achieves camouflage by producing and

giving off light that is directly proportional to the amount of ambient downwelling light for the purpose of counter- illumination.

Designers chose the pony fish as their source of inspiration. The function of camouflage now indicated the sub-function of producing a glow on the ventral side of the surfboard to match the ambient down welling light in order to prevent the formation of the silhouette. Now the issue became the mechanism of producing the light that achieved this function. In the case of pony fish, designers understood that the light is produced by bioluminescence – the light-producing organ of the fish houses luminescent bacteria Photobacterium leiognathi. This light is channeled from the light-producing organ to the ventral side and dispersed by creating rectangular light spots on the ventral side. Therefore, the function of producing ventral glow was decomposed in other sub-functions: produce light, channel and disperse light.

In order to produce light for the surfboard, the traditional means of having an onboard light source and a power source was considered an inferior solution. The search for alternate means of producing light sparked another round of search for biological sources of inspiration, which led them to an organism called Brittle star (a kind of a star fish). This organism implements the mechanism of photo-reception. The dorsal side of the Brittle star is covered with thousands of tiny eyes, or microscopic lenses, making the entire back of the creature into a compound eye. This mechanism can be used to collect surrounding light rather than having to produce luminescence as in Pony fish. This suggested a design in which the top of the surf- board would be covered with (suitably distributed) tiny lenses to collect the sunlight incident upon the surfboard.

In order to channel and disperse the light collected to the bottom, their design incorporated embedding optic fibers within the surfboard. One end of these cables would be connected to the lenses on the topside and the other end would be positioned on the bottom side. Although this would channel and disperse light, it would lead to spots of brighter and dimmer light when seen from below the surfboard. This would still produce

a silhouette, albeit of a different kind compared to the normal surfboard. To counter this, they had to think of another sub-function: disperse light to mimic the wavy pattern of the ocean surface. In order to achieve this function, their final design included adding a layer of "pattern light diffusers" on the bottom of the surfboard that disrupts the pattern of light (coming from the optical fibers) in controlled ways. This layer could be structured to mimic the wavy pattern of the ocean surface.

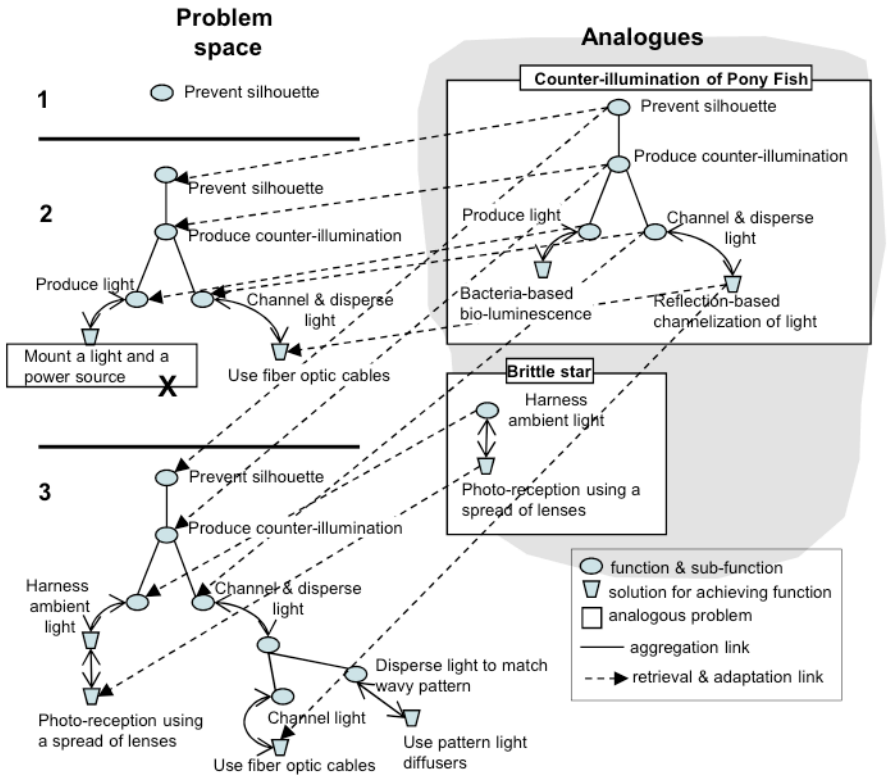


Figure 4.2: Design trajectory of Project InvisiBoard

Figure 4.2 depicts the evolution of this design solution. Step 1 depicts the nature of the problem space early in the design. The main function is the prevention of silhouette. Step 2 shows the retrieval of the pony fish analogue and the creation of two sub-functions: produce light, and channel and disperse light. For the first sub-function (produce light), Step 2 depicts the following: (i) solution in the source design (bio-luminescence) is not transferred, and (ii) the simple solution of mounting a light and

power source is rejected. For the second sub-function (channel and disperse light), a fiber optic-based solution is proposed in Step 2.

In Step 3, the search for a solution to the function of producing light has been transformed into "harness ambient light." We do not have a good explanation of this function transformation. A search based on this trans- formed function has led to the retrieval of the Brittle star analogue and the transfer of the photo-reception solution. Step 3 also depicts how the evaluation of partial solution of Step 2 has indicated that using fiber optic cables alone for both channeling and dispersing light does not eliminate the silhouette (but merely creates a different kind of silhouette). This has led to further decomposition of the original "channel and disperse light" function into two individual sub-functions. The channel light sub-function is still done through fiber-optic cables, but the dispersion is done through specialized "pattern light diffuser" devices. Knowledge about the diffuser devices was based on background domain knowledge and not gained by analogy as far as we can tell.

Case Report 3: Project Eye in the Sea

The goal of this project was to design an underwater microbot with locomotion modality that would ensure stealth. The problem was "biologized" as: "how do marine animals stalk their prey or avoid predators without being detected?" Two marine biological systems were considered as sources of inspiration, copepod and squid.

The initial research for the underwater microbot focused on the copepod as a source for understanding stealthy locomotion. In exploring this concept, designers became aware that the copepod used two rhythms (of leg-like appendage movement) for achieving motion underwater. A slow and stealthy rhythm was used during foraging for food, and a quick but non-stealthy rhythm was used during escaping from predators. This understanding led the designers to decompose their original problem into two separate functions, one for slow and stealthy movement, and one for rapid, yet stealthy movement.

Copepod acted as a source for generating a solution to the former part of the problem (slow and stealthy motion). While foraging for food, a copepod is not noticeable to its prey because it moves its appendages rhythmically in a way such as to minimize the wake produced in water. The knowledge of this mechanism, known as "metachronal beating pattern," was transferred from the copepod source to create a partial solution.

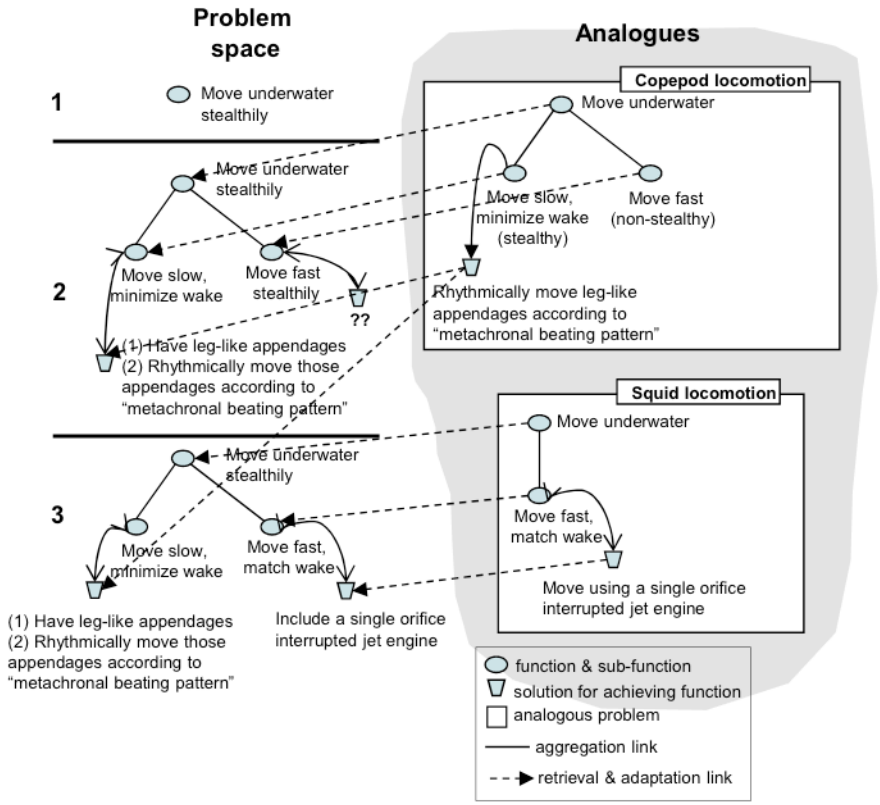


Figure 4.3: Design trajectory of Project Eye in the Sea

Next, the designers had to address the second sub-function (stealthy fast motion). They used the squid locomotion as an inspiration for achieving this function. Some squids implement a single orifice, interrupted, jet propulsion for forward motion. This mechanism simultaneously addresses two constraints. First, this kind of locomotion is much faster compared to the copepod's locomotion. Second, this kind of locomotion is stealthy because its wake matches the external disturbances that naturally occur in the

surrounding water. The stealth achieved here (wake matching) is significantly different from the way stealth is achieved in copepod motion (wake minimizing). Figure 4.3 visually depicts the generation of this solution.

Case Report 4: Project RoboHawk

The goal of this project was to conceptualize a bio-inspired bomb detection technology for chemical sensing and tracing of nitromethane and ammonium nitrate. The overall function of the bomb-detecting device was divided into two sub-functions, *motion* and *sensing*, which would interact such that the overall function is achieved. When designing the sensing device, the team was looking for a mechanism that would be mobile and would move freely within a designated area. The finished device should also be low-key. Presently, trained dogs are used for most bomb-sniffing operations in airports. These dogs are extremely conspicuous, due to the fact that people have come to associate trained sniffing dogs with bomb detection, or detection of other unwanted materials. The design team premised that by suspending the device from the ceiling, it will be kept out of the way of travelers, and will allow for unhindered maneuvering, and will provide the device a less prominent physical appearance. Due to the fact that the device will be suspended from the ceiling, it was optimal for the machine to move in a pattern that will allow it to maximize its coverage of the room where it will be located. The device would be detecting the presence of chemical compounds as it swept the room. When the device located a concentration of the desired chemicals in the air, it would relay this information back to a person who was monitoring the device. The device would specify its location so that security could be notified of the location of the bomb.

The most important function of the device was its ability to detect chemicals found in bombs. The chemicals that the device would be detecting were ammonium

nitrate and nitromethane, which were two common compounds found in homemade bombs. This device would recognize both components of the bomb, which would minimize the frequency of false positives. The design would also incorporate a computer that would function in a motor control and as an information relay system for the position of the RoboHawk as well as the presence of chemicals.

In the beginning, the design team looked to nature in an effort to biologize the challenge of detecting specific chemical components of homemade bombs. The components of the challenge included asking specific questions, and looking to nature to find the answers and alternate solutions to the questions. The team biologized the particular challenge of bomb detection as follows: What sort of signals do bombs give off and how do they compare to natures signals? What are examples in nature of organisms that must track signals and how do they work? How do organisms sense the chemical or signal?

One of the primary roles of the RoboHawk would be the chemical odor tracking of hazardous materials. In order to discover the most effective way in which to track an odor, an organism that performs this task particularly well would be studied. Antarctic procellariiform seabirds were chosen because they are exceptionally good at tracking scents over an extreme distance.

Initially, it was predicted that these birds would sense odors as they travel and remember the locations and relative strengths of target odors in a sort of mapping technique. Studies had shown this to be inaccurate. In order to find their prey, these animals simply traveled until they sensed a strong enough odor plume that would indicate a significant presence of prey. Once they reached this area, they then changed their behavior in order to narrow down the source of the target odor. Most seabirds would fly directly crosswind in order to find a strong scent. This flight pattern maximizes the chance of the bird detecting an odor plume as it travels intermittently in turbulent eddies. But once a more significant odor plume is detected, the seabird would then fly upwind in

a zigzag pattern. This model was thought to maximize the chance of detection and also help to narrow down the location.

It was noted that not all birds conform exactly to this behavior. For example, the albatross flies in a more oblique pattern relative to the crosswind. It was speculated that this flight is due to energy conservation combined with a tracking approach that may be based more on visual than olfactory sensing. The design team's device would not be hindered by fatigue, and it was not be able to rely on sight to track its target. For these reasons, the principle of odor tracking minus the fatigue compensation was carried over into the design. This concept of tracking specialized plumes was considered very useful in the tracking of dangerous materials such as bombs. The zigzag algorithm was used as a tactic to employ in automated devices used in odor sensing. Though the seabirds' tracking ability was complex and highly sensitive, it did encompass a large-scale operation. Therefore, RoboHawk would have the ability to operate in a large workspace if needed.

The final design of RoboHawk had the following features. In order to address the sub-function of motion, i.e., moving in the direction of a suspected signal, an approximation of the path following algorithm of sea birds was used as the source of inspiration. Due to this RoboHawk did not move in a straight-line path towards the signal, but utilizes a particular zigzag path-following approach for target cuing in order to maximize their search. In order to address the second sub-function of sensing, it was understood that two types of chemicals needed to be detected: ammonium nitrate, and nitromethane. In order to detect ammonitrate, the sensor would use the Surface Acoustic Wave (SAW) mechanism. This approach was not biologically inspired. On the other hand, in order to detect nitromethane, the sensor would use the mechanism of Membrane/Enzyme system found in the olfactory organs of many critters like moths, roaches, dogs, etc. This approach was biologically inspired. It is not clear from the data

as to why they used bio-inspired mechanism for detecting one substance and not for the other.

The following sections outline the key findings of this study. Only those findings that are relevant to this thesis and which pertain to the task of bio-inspiration seeking are presented here. A full list of findings can be found in other published sources (Helms, Vattam & Goel 2009; Vattam, Helms & Goel 2008; Vattam, Helms & Goel 2007).

Multiple Analogies

The four case reports above suggest that a single design solution may sometimes require multiple biological sources of inspiration. One single biological source was not always available to help solve a target problem in its entirety. In such cases, the design solution was generated in a piecemeal or modular fashion by composing the resultants of multiple analogies to smaller sub-problems, while the target design problem evolved with each application of an analogy. This implies that the act of actively seeking bio-inspiration is not a one-time exercise in a given design episode. In the current study, 6 out of 9 (66%) design solutions were generated by composing the results of multiple cross-domain analogies. Similarly, in the following year (Fall 2007), 4 out of 10 (40%) of the design solutions were compound solutions.

One implication of such cases of compound biologically inspired design situations is that a single design episode requires multiple undertakings of the task of bio-inspiration seeking. Dues to this aspect, any improvements in giving support to designers accomplish this task can have a multiplier effect, resulting in greater payoff for designers.

How were the Sources of Inspiration Obtained?

From what I observed in the classroom, designers used three basic approaches to find their biological sources of inspiration: serendipity, human-human interaction, and human-information interaction.

The serendipity approach refers to the biological systems that designers "found" as a result of the found object exercises. Classroom activities included regular found object exercises that required designers to bring in biological samples that they encountered in their everyday lives and analyze the solutions employed by these samples. These exercises were intended to expand awareness of biology, provide hands on experience with biological systems, and encourage the designers to dig progressively deeper into the functions of biological systems. Additionally, journal entries required designers to write about their classroom experiences, including found object discussions, and to document their own design thinking.

The human-human-interaction approach to bio-inspiration seeking involved interacting with other people and eliciting knowledge about biological systems that might potentially be useful for solving a team's design problem at hand. These social interactions were: 1) intra team, directed towards the biologist in the team; 2) inter team, where one team learnt from the exploration and experiences of peers in other teams; 3) external, where team members interacted with either the instructors or other domain experts (faculty mentors, guest researchers) in order to learn about potential biological systems to look into.

The human-information-interaction approach involved searching online for biological information about systems that are analogous to the target design. Based on my observations I suspect that this was one of the predominant approaches for finding biological sources of information. I also suspect that this activity was carried out individually by design team members, as well as collaboratively where they coordinated their information seeking activities.

With respect to human-information interaction approach, designers reported using a range of online information environments to seek information resources about biological systems. These included: 1) online information environments that provided access to

scholarly biology articles like Web of Science, Google Scholar, ScienceDirect, etc., 2) online encyclopedic websites like Wikipedia, 3) popular life sciences blog sites like Biology Blog, 4) biomimicry databases like AskNature, and 5) general web search engines like Google. But the most frequently used environments were the ones that provided access to scholarly biology literature like Web of Science and Google Scholar. Biology articles, both scholarly and otherwise, were the predominant types of media or information resources consumed during the process of online bio-inspiration seeking.

Strategies for cross-domain information seeking

The target problem and the source of inspiration are situated in different domains. If cues from the target problem alone are employed during search, then designers are likely to find information resources that belong to the same domain as the target rather than biology. I observed that instructors suggested several strategies in order to bridge the engineering-biology divide.

The first strategy was to "biologize" the problem. Biologizing the problem involved redefining the problem by taking the key concepts in the design problem and substituting them with similar biological concepts. Then the concepts from the biologized problem were used as cues in order to search for biological systems. For instance, in the BriteView project, the concept of a light-emitting material that resisted drowned illumination in sunlight was biologized to organisms producing iridescent colors in the presence of sunlight. Then the concept of iridescence was used to find biological systems that had this feature. Although this process of biologzing the problem was observed in all the design projects, it remains a black box: there were no explicit rules for how to do this, but relied on the tacit skills of designers. This seems to be an effective strategy because

all the teams reported the biologized problems in their project reports. The other strategies that instructors mentioned are included in Table 4.1.

Table 4.1: Strategies for cross-domain information search

Search strategy	Strategy description
Change constraints	If the problem is narrowly defined, such as 'keeping cool', change the constraints to increase the search space, for instance to 'thermoregulation'.
Champion adapters	Find an organism or a system that survives in the most extreme case of the problem being explored. For instance, for 'keeping cool', look for animals that survive in dessert or equatorial climates.
Variation within a solution family	Where multiple organisms have faced and solved the same problem in slightly different ways, e.g. bat ears and echolocation, look at the small differences in the solutions and identify correlating differences in the problem space.
Multi-functionality	Find organisms or systems with single solutions that solve multiple problems simultaneously.
Inverse functions	If a particular function is not yielding many biological solutions, inverse the function. For instance, if the function is 'keeping cool,' look for organisms that achieve the function 'keeping warm.' In some cases, the inversed function might yield many potential systems. Learning about the mechanism for the inverse function can sometimes yield insights into accomplishing the original function.

It should be noted that the core of these strategies involve using certain abstractions to bridge the domain of technology and the domain of biology. These abstractions include *functions* (e.g., strategies like biologizing, inverse functions, multi-functionality involves

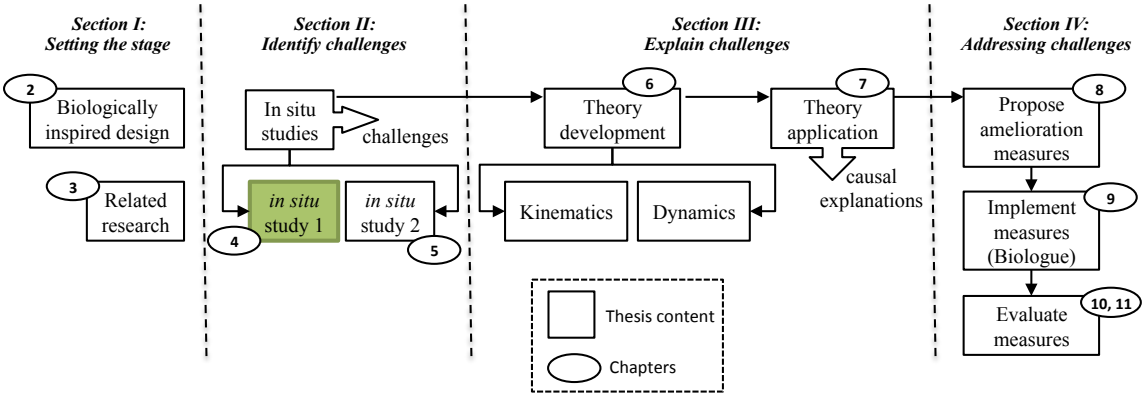
abstracting and/or transforming the functions), *operating environment* (e.g., the strategy of champion adapters), *mechanisms*/*physical principles* (e.g., the strategy of variation within solution family involves similarity across mechanisms and principles), and *constraints* (e.g., the strategy of changing constraints).

Online bio-inspiration seeking experience

My observations of designer-designer interactions and designer-instructor interactions indicate that the online information environments on which designers relied upon did not adequately support the task of online bio-inspiration seeking. I noted that it took a long time for designers to find their biological sources of inspiration using the online approach (several weeks of searching). Designers complained that the information-seeking process was frustrating because the search process consumed a lot of time but yielded very few articles containing biological systems that were actually useful in addressing their target problem. I also observed that although designers spent a lot on time searching for novel material, in many cases they ended up using biological systems that they were already exposed to during their class lectures because their search process did not yield any new sources. This issue of difficulty of online bio-inspiration seeking was exasperated due to the fact that, in many cases, designers had to undertake this task multiple times in the course of their design episode due to compound analogies.

One of the limitations of this study was that we were not able to identify the specific problems that designers faced during online bio-inspiration seeking. This was because we could not observe them "live" as they under took this task outside the scope of our observable contexts (e.g., home).

Chapter Summary



This chapter presents an *in situ* study in which I have analyzed the biologically inspired design process in terms of the practices of the designers and their products. A number of insights related to the task of bio-inspiration seeking were gained in the process. First, we noted that bio-inspiration seeking is characterized by the search for one or more cross-domain analogies between the target technology under design and source biological systems. Second, we noted that there are multiple methods for obtaining biological sources of inspiration. But one of the predominant ways in which designers obtain their biological sources is by means of human-information interaction, i.e., by going online and searching for articles that contain biological systems that are analogous to the target design. Second, we noted that designers use a range of different online information environments to obtain their sources. But the most widely used ones were those which give them access to scholarly biology articles like Google Scholar, Web of Science, etc. Third, we noted that designers use certain strategies like "biologizing the problem" and employ key abstractions like functions, mechanisms, principles, etc. in order to cope with the cross-domain nature of the task. Fourth, in spite of such measures, we noted that it took a long time for designers to find their biological sources of inspiration using the online approach. Designers complained that the information-seeking process was frustrating because the search process consumed a lot of time but yielded very few articles containing biological systems that were actually useful in addressing their target problem. From these observations, it can be inferred that although designers

rely on certain online information environments to obtain their sources of inspiration, those environments do not adequately support the task of online bio-inspiration seeking. Therefore the task becomes very inefficient and ineffective when carried out in conventional online environments. Fifth, we also noted that a single design episode sometimes result in multiple undertakings of the bio-inspiration seeking task due to compound analogies, which magnifies the above issue. On the other hand, if we can find ways to improve the efficiency of their information-seeking process, it can have a multiplier effect in this context.

Although the study presented in this chapter indicates that designers were having trouble with online bio-inspiration seeking process, we were not privy to the specific nature of the problems they encountered. This is because we were not able to observe the task "live" due to the fact that designers undertook this task in outside the contexts we were able to observe (e.g., home). In the next study presented in the next chapter, I became a participant observer rather than an outside observer. This gave me firsthand access to the task and an opportunity to identify the specific challenges associated with the online bio-inspiration seeking task.

CHAPTER 4

A FOLLOW-UP STUDY OF ONLINE BIO-INSPIRATION SEEKING

This chapter presents my second *in situ* study of biologically inspired design. The objective of this study was to gain a better understanding of the processes of online bio-inspiration seeking process, including the specific nature of challenges that designers face when engaged in this task. This study was again conducted in the context of the same undergraduate interdisciplinary course on biologically inspired design at Georgia Tech (ME/ISyE/MSE/PTFe/BIOL 4803), but offered in a different year.

In 2008, this course attracted 43 students. The class was composed of 16 biologists, 2 biomedical engineers, 10 mechanical engineers, 7 industrial engineers, and 6 material science engineers, and 2 computer science engineers. Similar to the previous study context, the students were grouped into 4-5 member design teams, with at least one biologist in every team, to work on their semester-long biologically inspired design project. In terms of student demographics and the course structure, there was not much change from the Fall 2006 version of the course.

Participatory Study

The approach adopted in this study is *participatory research* (Reilly 2010). In the first study we were not able to directly observe the students' online bio-inspiration seeking behavior because much of it happened outside the classroom and design meetings, settings that we were not able to observe. In order to overcome this issue, I decided to register for the course and participate in the design process. With full participation, I would be an integral part of a design team and would not only get firsthand experience with the bio-inspiration seeking task, but also a chance to closely interact with and understand the experiences of fellow team members who were also engaged in this task.

Participatory research is research that is fully collaborative and emphasizes deliberate participation and contribution of the researcher in the actively examining some issue related to a community of practice. The following guiding principles of participatory research were observed in this study. (1) Ensure meaningful participation by the researcher such that the "insiders" (participants) and the "outsider" (researcher) are all partners invested in the project results. (2) There is an assumption of co-equal status of the practitioner and the researcher. (3) The formal boundaries between traditional roles (researcher-subject/participant) are reduced to reduce special status accorded to the researcher. (4) In participatory framework, objectivity is not the gold standard; rather critical subjectivity and reflexivity are valued.

In this second study, I enrolled in the course for credit as any other participant would. I engaged in all the academic activities that were required as part of this course. Recall that this is a project-based course. When the design teams were formed during the initial stages of the course, I became part of a design team called FORO. This study focuses on the design activities of my design team and what I learnt as a productive member of this team.

Data

In terms of data, I maintained a field note journal where I noted my observations. My notes included observations of: 1) across the board student-instructor and student-student interactions inside the classroom, 2) team FORO interactions inside the classroom, and 3) team FORO interactions outside the classroom, mostly restricted to design team meetings. With regard to team FORO interactions, I recorded the thoughts expressed by members during the team meetings, the concepts that were discussed, the ideas that were thrashed out, the external representations that were constructed (e.g., diagrams on the white boards), etc. Additionally, I noted my own information seeking experiences like the keywords that I used, the results that information environments

threw at me, the problems that I faced, how long it took me to find information sources, etc. Then I discussed some of these issues that I faced with my team members and noted their thoughts on those issues in my field journal as well. All electronic and non-electronic communication between team FORO members was also used as a data point for this research.

Each team member also maintained his or her own idea journal. This was similar to what participants did in the first study. It was a course requirement. This was a free-form journal in which every individual designer in the course was required to not only externalized their design thoughts and the biological systems they encountered, but also reflected on the design process and the activities they were undertaking. In order to motivate students to make effective use of the journal, it was collected and evaluated by the instructor every week from a small set of randomly selected students. My team members' and my own idea journal was another source of data for this study.

I also collected all the work products generated by team FORO, which was another data point for this study. These were usually milestone documents. After every milestone (e.g., problem definition, biological search, initial design, or design analysis), we were required to submit a document detailing the team's accomplishments leading up to that milestone. These documents contained snapshots in the progression of the design over the semester-long duration. The final design report, which captured different aspects of their finished design, also provided another data point for my analysis.

Upon joining this team, the other team members were made known that in addition to being a team member I was also a researcher interested in studying the design practices of the team. It was made known that my observations, notes of team conversations, team's electronic and non-electronic communication, etc. would be used as data for my research. It was made clear that my role as a researcher would not directly influence or reduce my participation as a team member. It was also made clear that the

data that I collected would not be shared directly with the instructors during the term, and that my researcher role would not influence the their grades in any way.

But there were two major points of departure from the first study with respect to data collection and analysis. The first departure pertains to the cross-case synthesis. In the first study, there were multiple design teams that were studied. Therefore we had more than one case to reflect upon, which made cross-case synthesis possible. In contrast, this study follows the activities of one team alone. Therefore, only one case description is developed and the findings are specific to this one particular team, which can be contrasted with the compilation of observations that are common across multiple teams in the first study.

The second major departure pertains to the involvement of only one researcher. In the first study, the findings were the result of the analysis of two researchers who collected the data independently of each other. In contrast, the analysis in this study is based on the observations and reflections of one researcher alone.

Case Report: Project FORO

We begin with the case description of the design project attempted by team FORO. The design challenge for this team was to conceptualize an energy-efficient desalination technology for converting seawater to potable water. This team consisted of 6 team members: 2 mechanical engineers, 1 biologist, 1 industrial engineer and 2 computer science engineers.

The team Formation Process

The design team formation process began in the third week of the course. In Fall 2008, this course had a design theme around which the design teams were required to identify problems and propose solutions to. The design theme that motivated the design problems this year was the issue of *global water crisis*. The process began with the

viewing of a documentary. Students attended the screening of an award-wining documentary called “Flow” which was arranged by the instructors in a Georgia Tech auditorium. *Flow* documents the global water crisis, from the role of corporations in global water supplies, to how we are quickly running out of water, ultimately asking the question “Can Anyone Really Own Water?” The documentary contains interviews with scientists and community activists, as well as beautifully captures the role that water plays in each of our lives.

Based on this documentary, every student was required to document an issue/problem that they liked to address. These were collected by instructors and organized into different problem areas like alternative sources of drinking water, water harvesting, reducing water consumption, preventing water wastage, etc. Students in each category then signed on to specific projects in each problem area. Students were then shuffled around within the problem areas such that each team composition reflected diversity in backgrounds, making sure that each team was assigned at least one biologist. I signed on to one of the teams that were interested in finding alternative sources of drinking water. This team was later named team FORO.

Problem Definition and Elaboration

Team FORO decided to address the problem of increasing water shortage on a global scale by designing a novel water desalination technology that converted ocean water into a drinkable supply of fresh water. Initially, we surveyed five existing desalination technologies. Three among the five were thermal based processes (multi-stage flash evaporation, multi-effect distillation and vapor compressed distillation). Two of them were membrane-based processes (reverse osmosis and electrodialysis). In the course of our survey we learnt that current desalination technologies employed processes that were very energy intensive, which prevented their widespread adoption. Therefore

we added a new constraint to our design problem: our solution should use significantly less energy compared to the existing technologies.

Online information seeking played a central role in the survey. The *function* of desalination was used as a cue to retrieve existing industry-standard desalination technologies. This survey served two purposes. First, the different sources in our survey helped us infer different mechanisms (or physical processes) used for achieving the function of desalination. Second, the different sources helped us to elaborate our problem by suggesting alternate problem decompositions. These decompositions were related to each other through a hierarchy of functions that would lead us towards our design goal, producing a kind of a problem elaboration schema. Problem decomposition requires knowledge of the form $D \rightarrow D_1, D_2, \ldots, D_n$, where D is a given design problem, and D_is are smaller sub-problems. In many instances, this knowledge was inferred from the design patterns abstracted from the current technologies surveyed. By design patterns I mean shared generic abstractions among a class of designed systems. For instance, all membrane-based desalination technologies share common functions, mechanisms and principles.

Evidence for these design patterns comes from diagrams like the one shown in Figure 5.1. This diagram was reproduced here from team FORO's design report. The evidence for the problem elaboration schema - a higher-level knowledge structure that relates design patterns and other abstractions to each other - also comes from a diagram shown in Figure 5.2, again reported in the team's problem definition document. These diagrams were reconstructed from the diagrams that were jointly constructed by team members on white boards in the design meetings.

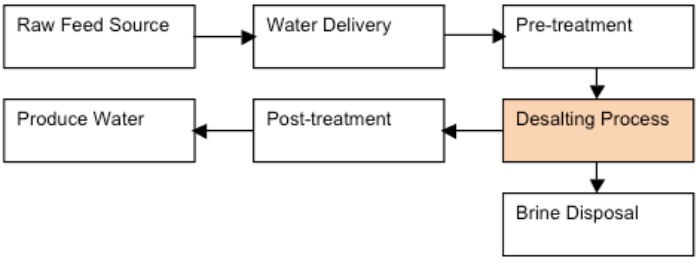


Figure 5.1: A design pattern for membrane-based processes

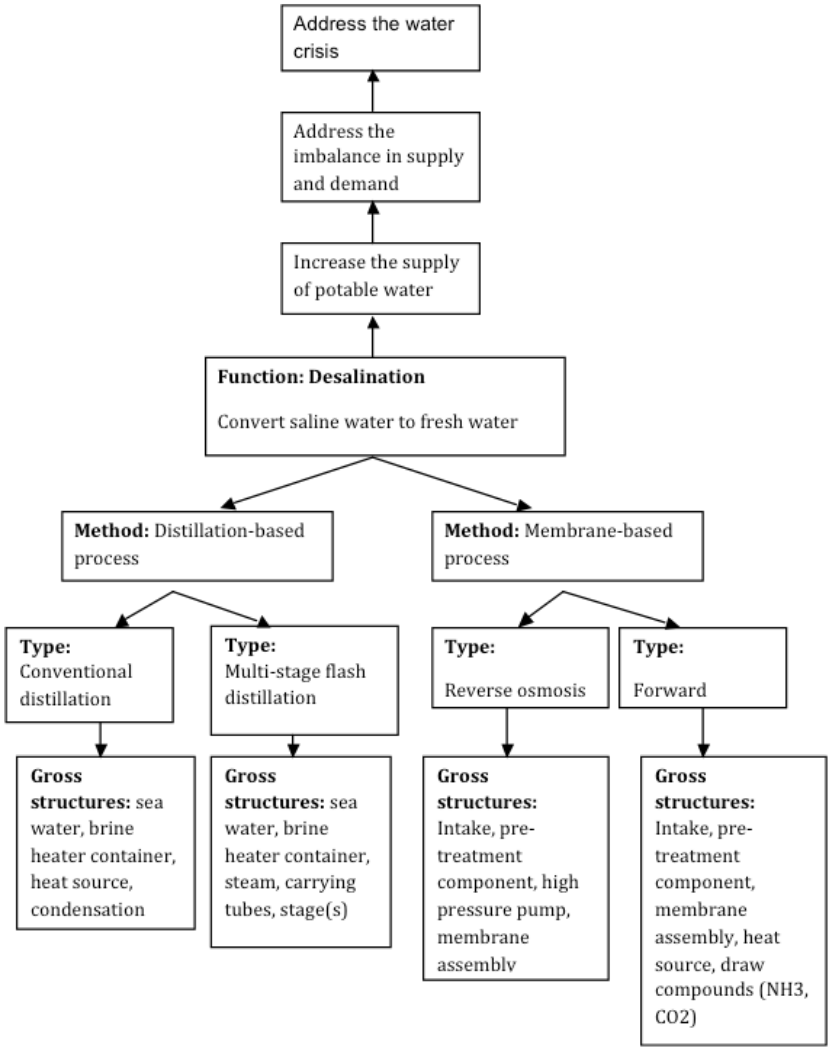


Figure 5.2: A problem elaboration external representation

Search for Biological Sources of Inspiration

We used our developing knowledge of the desalination problem to find biological analogues that were applicable to our problem. This search was a collaborative process and predominantly centered around looking for scholarly articles that discussed biological systems that could help solve the desalination problem. Although we individually looked for biology articles on our own, we coordinated and monitored each other's search activities and shared the results. The problem elaboration schema from earlier activity provided the reference point for our search process. Paying attention to different aspects of the problem elaboration provided different cues for the retrieval process. A total of 24 biological systems were identified by the team at various stages of this biological exploration activity, which spanned more than one third of the semester. However, around ten systems were given serious consideration: supra orbital salt glands in penguins, salt glands in marine reptiles, gills in salmons, respiratory tract in camels, kidneys, root systems in mangroves, esophagus in Gobius Niger fish, esophagus in eels, aquaporins, small intestines in humans and other animals. Drawing analogy between our problem and the identified biological systems helped us infer different mechanisms for achieving a desired design goal.

Three different strategies of retrieval were noted here. First, functional cues from the elaborated problem were directly used to retrieve biological sources. For instance the function of desalination or the related "removal of salt" was used to retrieve sources like supra orbital salt glands in penguins, salt glands in marine reptiles, gills in salmons, etc.

Second, the general abstractions in the problem elaboration, like the aforementioned design patterns, were used to retrieve biological sources. This explains how a certain source like the *small intestine* was retrieved when there was no reference to salt anywhere in the intestine process (the intestine source included sugar solutions and not salt solutions).

Third, design patterns were sometimes transformed and those transformed patterns were then used to retrieve biological sources. This provides one explanation for the retrieval of the *camel* analogue to the thermal desalination process. The function of camel's respiratory tract is to (1) saturate and warm the inhaled air so that it is suitable for the lungs to process and (2) desaturate and cool the exhaled air so that the moisture and heat are conserved and are not lost to the environment. This system, which had no relation to concepts like desalination or salt or solutions or energy expenditure, was still suggested to as an analogy to the thermal desalination process. This can be explained by the transformation of the design pattern for thermal process shown in Figure 5.3a (seen from the perspective of what is happening to the water) to a pattern shown in Figure 5.3b (seen from perspective from what is happening to the air surrounding the water) and by comparing the camel's case to transformed pattern.

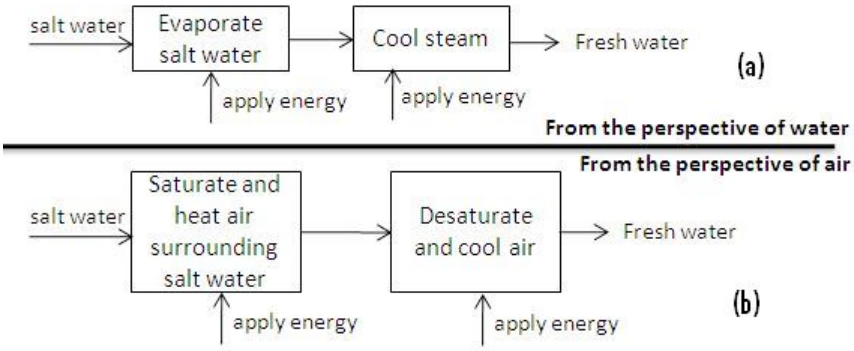


Figure 5.3: Pattern transformation

This bio-inspiration seeking aspect of the design process will be elaborated further in a subsequent section on key observations towards the end of this chapter.

Initial design development

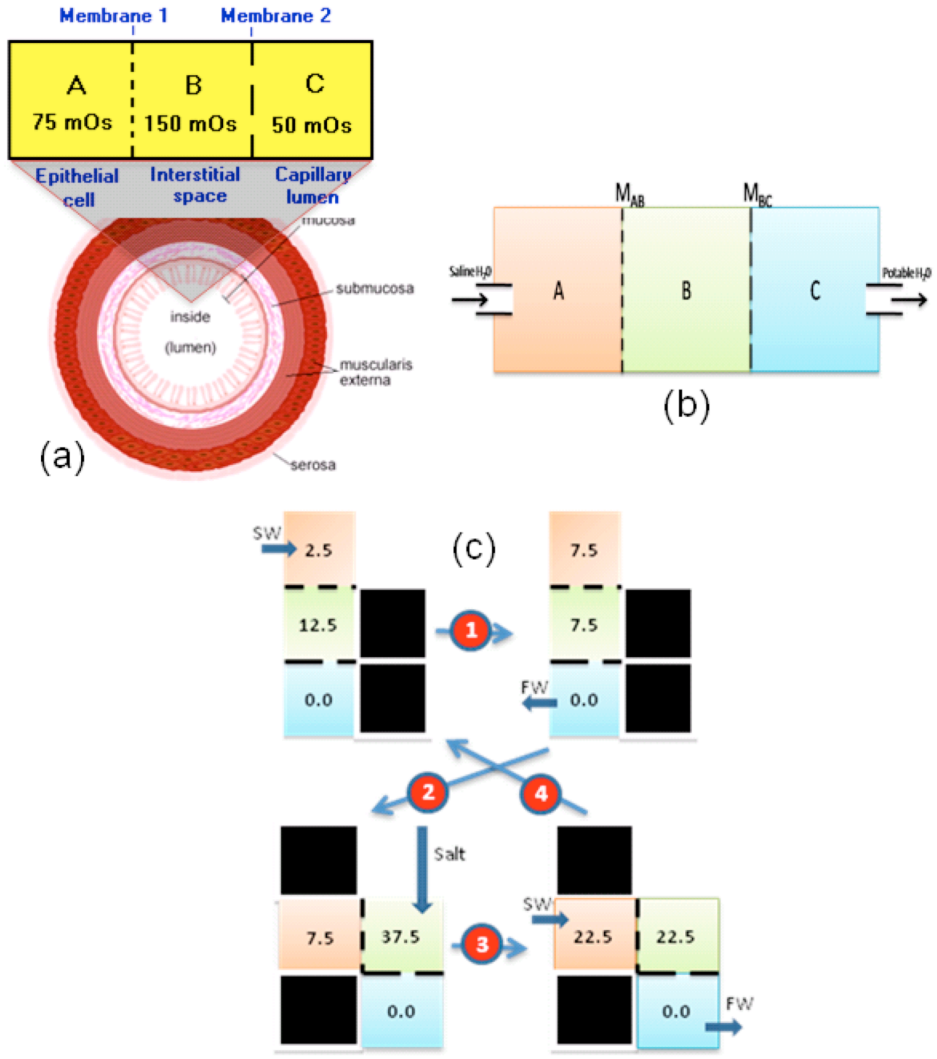


Figure 5.4: (a) biological source (intestine); (b) initial design solution; (c) redesigned solution

Developing a biologically inspired design solution involves retrieving a suitable biological system, understanding how that system works to a sufficient degree of depth, extracting mechanisms and principles associated with that system into a solution-neutral form, and applying those mechanisms and principles in the target domain of engineering. Team FORO had identified a subset of promising biological analogues. These systems were understood by the designers to varying degrees of depth. Based on our understanding, those systems were classified as using active transport (requiring external energy in the form of ATP) or not. This classification was used as an elimination criterion

- biological systems that used active transport were deemed unfavorable (because the goal was to achieve desalination with minimal energy expenditure). This eliminated all sources but the small intestine, camel nose and mangrove roots. Not enough was understood about the mangrove roots, and it was not readily apparent how the camel nose mechanism could be implemented as a solution. Therefore, team FORO developed an initial design solution based on the mechanism of the small intestine.

The small intestine reabsorbs water using a conjunction of forward- and reverse-osmosis principles, a method called the three-chamber method. This mechanism was transferred to the target problem to produce an initial design solution. Figure 5.4a and 5.4b above shows a side-by-side comparison of the biological source and the initial solution developed.

Design evaluation

Team FORO now had produced a conceptual design of a desalination technology that was not only novel, but also eliminated the need for applying external energy (except for the energy required to feed the ocean water), which was too good to be true. We took our solution to an expert with several years of research experience in membrane technology for evaluation. The expert suggested that our initial design would not work. This was because the flow of fresh water in our design depended on maintaining the salt concentration gradients in the three chambers. But our design worked in such a manner that the salt concentrations in each chamber would change over time to offset the gradient, reaching equilibrium and stopping the flow of water.

The expert came to this conclusion with the help of an analogy. He saw our initial design as a device that contained a piston pushing liquid from one end of a cylinder to the other, with a semi-permeable membrane attached to its far end. The flow is maintained in this device as long as one is applying force on the piston. The reaching of the equilibrium in our design was akin to someone taking his or her hands off of the piston. The cognitive

purpose of the expert's analogy in this case was to evaluate the design and identify any potential problem in it.

Redesign

Now the new challenge for team FORO was to redesign our system such that it did not reach equilibrium. We redesigned our system by coupling two three-chamber systems and by configuring those two to work cyclically. When the first three-chamber system reached equilibrium, it would create non-equilibrium conditions in the second three-chamber system, ensuring that the water would continue to flow from the second one, and vice versa. The redesigned system is depicted in Figure 5.4c. This is again reproduced from the team's design report.

Design analysis

The next step for team FORO was to perform a quantitative analysis of our design in terms of estimating the flow rate of the fresh water produced. If the flow rate was of the order of cubic centimeters/hour, as was the case with the intestine, then our design was not viable. We had to determine how well the designed system scaled up compared to its biological counterpart. Since the biological model did not contain a flow analysis, the equations had to be derived from first principles. None of the team members knew fluid mechanics well enough and had to rely on the expert for our analysis guidelines. But at that time, our expert mentor was traveling and was not available. So we put their analysis on hold till we could find another expert who was willing to help us.

A few days later, one of the designers came across a paper by Popper et al. (1968) by chance. This paper presented a novel mechanical system for chemico-mechanical separation of solutes form a solution. This system was both similar to and different from our design. Popper's system was similar because it used forward-osmosis in conjunction with reverse-osmosis to achieve desalination. At the same time, it was different because

its structure was different and did not utilize a three-chamber method. This system was also prone to reaching a steady state resulting in the stoppage of flow, and was not biologically inspired. However, Popper's paper had a flow analysis of that mechanical system. Recognizing that Popper's mechanical system was analogous to our design, we adapted and transferred the flow equations from Popper's situation to our current design specifications, estimating a peak flow performance of 139.967 l/hr. With the successful completion of the quantitative analysis, our design episode came to a conclusion.

Key Findings

This section outlines some of the key observations specifically related to understanding the nature of online bio-inspiration seeking process and the challenges associated with it.

Collaborative nature of online bio-inspiration seeking

The observed process of online bio-inspiration seeking was *collaborative* in nature, characterized by a shared goal and joint effort in accomplishing the goal. Online search is generally treated as a solitary activity, with Web browsers and search engines typically designed to support a single user working alone. However, in this context, designers in a team worked together to accomplish the task of bio-inspiration seeking.

Computer supported collaboration can be classified along the dimensions of time and space as synchronous or asynchronous, and co-located or distributed respectively (Dix 1998, p. 465). The nature of collaboration during online bio-inspiration seeking in our observations was asynchronous and distributed. During much of the process, designers worked independently and in their own separate spaces. But as opportunities presented themselves, individual members would share their findings, solicit help, clarification or feedback electronically via email. For instance, team members would

share an article that they thought was relevant via email, as seen in the following two examples:

---

date: Tue, Oct 28, 2008 at 6:54 PM
subject: bio articles for thursday

Hey Desalinators,
…
…
Here's one really good article from web of science: http://proquest.umi.com.www.library.gatech.edu:2048/pqdlink?Ver=1&Exp=10-27-2013&FMT=7&DID=1564294941&RQT=309&clientId=30287. This article is about inland desalination and supplementing existing supplies or-to control the salinity of current sources with zero liquid discharge (ZLD) approach.
…

---

date: Tue, Nov 18, 2008 at 4:39 PM
subject: osmosis and thermodynamics

Team,

Thought that this might be a useful paper on the thermodynamics of osmosis (concise, recent and from a prestigious journal).

The following is an example of team members exchanging email to seek help/clarification:

---

date: Tue, Nov 25, 2008 at 6:36 PM
subject: Simple Animations to help you guys understand pumps

Hey guys

I found these for you... It's short (like 1-2 mins) and very simple to follow so I hope you guys all take a look at it

1. sodium - potassium pump
[http://highered.mcgrawhill.com/sites/…]
2. proton pump
[http://highered.mcgrawhill.com/olcweb/cgi/…]
3. cotransport
[http://highered.mcgrawhill.com/olcweb/cgi/pluginpop…]

On Tue, Nov 25, 2008 at 3:01 PM, [ ] wrote:

Is active transport the same as cotransport? Does anyone know

---

Additionally, I observed that the whole team would periodically meet face-to-face to coordinate their activities. These activities included the following.

Division of labor: team members jointly identified areas of information need and each member was assigned an area need to fulfill. For example, *"how do organisms adapted to seawater environment deal with excess salt in their bodies"* is one example of an area of information need that was assigned to one of the team members. Elaboration of this aspect and examples are supplied further down in the discussion.

Sharing of knowledge: team members shared what they learnt as a result of their research. This mostly pertained to the workings of identified biological systems, how it is applicable (or inapplicable) to the design problem at hand.

Negotiation: team members negotiated with each other on various different issues, including division of labor (what areas to research, what sections to write in case of written deliverables, what sections to present in case of oral presentations, etc.), problem definition (scoping the problem, what requirement to include or exclude, defining performance characteristics, etc.), and selection of biological source of inspiration.

Joint problem solving: there were instances where the entire team came together in order address a pressing issue and make progress. For instance, when flaws were detected in team FORO's initial design (refer to the case study description above, esp. design evaluation and redesign sections), the team redesigned their solution in one of the team meetings through intense discussions and brainstorming.

Duration of the task

While typical online information-seeking acts are accomplished by employing a few queries in one session, the online bio-inspiration seeking process was much more *exploratory* in nature, extending over a significantly longer period of time and over multiple sessions. In the case of team FORO, the whole bio-inspiration seeking process lasted for a little more than seven weeks. In those seven weeks, individual team members

reported spending, on an average, 2 to 3 hours every week on the information seeking acts alone.

Initial stages of the process

The initial stages of the online bio-inspiration seeking process dealt with *defining the target* problem and establishing a common understanding of the design requirements. This pre-search stage was mostly social in nature, characterized by face-to-face communication and negotiation among team members. This process was facilitated by *functional decomposition* and *functional optimization* exercises:

Functional Decomposition: As problem definition was carried out, initially simple-seeming problems became complex, often involving multiple, integrated functions. Functional decomposition exercise is a conceptual tool that helps take a complex function and decomposes it into sub-functions. In the case of team FORO, a number of sub-functions were identified: source raw water feed, water delivery, pre-treatment, desalting, post-treatment, brine disposal, and device cleaning and maintenance. This allowed the team to negotiate on focus on functions that were most important to the project. In this case, the desalting function was made the central focus of the project.

Functional Optimization: Functional optimization exercise defines a function or set of functions in terms of an optimization problem or equation. This has upstream activity has several uses downstream in the design process. Abstracted to this level, designers can more easily transfer engineering requirements to biological solutions (and vice versa). Designers can also analyze potential new solutions by measuring performance against optimization criteria. In the case of team FORO, optimizing the energy required to desalinate the source water was made the central focus.

Both these techniques were conceptual tools that were taught as part of the course and mandated by instructors to incorporate into the design process. Team FORO dedicated one team meeting to performing functional decomposition and optimization

exercises. At the end of this phase, the team came to a common understanding of the problem that they were going to address:

"Our objective in this project is to engineer a bio-inspired desalination (specifically desalting) process that takes one unit of feed water as input to produce an equivalent unit of fresh water. The constraints for our deign are: (1) the TDS of the output fresh water should be fit for human consumption, specifically drinking (which is less than 500ppm), (2) the energy requirements for the achieving the desalination should be less than the most efficient among the existing techniques, preferably approaching the energy requirements of obtaining fresh water by pumping ground water. We make the following simplifying assumptions for our design. (a) The feed water (ocean water) is already filtered and pre-treated to remove all other unwanted contents. We only deal with pure saline water. (b) The total dissolved solid (TDS) content of the feed water, a measure of its salinity, is between 35000 and 50000 ppm. (c) Our design will not actively control for other parameters like pH, alkalinity, free residual chlorine, boron, total hardness, etc., which are all valid parameters that needed to be considered for any realistic implementation of desalination technology."

Once the problem was defined, the team had to *biologize* the problem as a precursor to the search process. During one team meeting, team FORO members negotiated several versions of the biologization of the problem. Four broad questions were agreed upon as potentially good directions for the search to take place: How do organisms adapted to living in seawater deal with excess salt in their environment? How do organisms that move between seawater and fresh water deal with the change in salt concentrations? How do kidneys filter blood to remove dissolved waste substances in order to maintain homeostasis? What biological processes result in evaporation of water?

These four biologized questions formed the basis for the *division of labor* for the subsequent bio-inspiration seeking process. Two team members decided to research the organisms for question 1, two members for question 2, one member for question 3 and one for question 4. While Questions 1, 2 and 4 were more open ended (specific biological systems were needed to be identified), Question 3 already assumed that biological systems like kidneys were a potential source of inspiration and needed to be understood in greater detail. Once the different information needs were identified and distributed, team members worked on fulfilling those needs relatively independently. This phase involved each team member searching for and making sense of information materials in online information environments in relative isolation.

The online search part of the process

After the division of labor, team members went off on their own to fulfill the assumed responsibility of finding biology articles. The various different information environments that played a part in the online bio-inspiration seeking process will be referred to here as the *information ecology* of the process. Based on my interactions with other team members and on my own experience, I noted that team FORO's information ecology consisted of: Google, Google Scholar, Web of Science, Wikipedia, ScienceDirect, AskNature, BIOSIS previews, PubMed, Biology-blog, Nature.com, Compendex, Knovel library, Bioblog, Science Daily, Youtube, TED talks, Biomimicry Institute website. This list was compiled by asking each tem member to self report all the online environments they remember visiting.

Furthermore, I also noted that there were one or more *preferred* information environments in this ecology, which were both individual specific and context specific. Team FORO members self reported that Google, Google Scholar, and Web of Science were the most heavily used online resources, followed by Wikipedia and AskNature. In my case, I generally began my search within Web of Science. Google Scholar was my

second choice. The reason I preferred Web of Science was because I could filter the search results by discipline. Since most of the time I was looking for biology articles, Web of Science allowed me to select only biology articles. I could also further filter the results by sub-disciplines like zoology, ecology, etc. This feature, which was missing in Google Scholar, was very useful to cope with the information overload problem.

If the chosen information environment did not yield desired results, it was common to switch from one information environment to another. This switching also occurred if the information needs changed due to what has transpired in the search up to that point. For instance, one may need to look up the meaning of a particular concept that was encountered during search. Wikipedia is more suitable for fulfilling this need as opposed to, say, Google Scholar.

From my own search experience, I noted that within particular information environments in this ecology, the individual information-seeking process in the context of BID was not very different from other everyday information seeking tasks and included a range of common search and browse tasks, including some mix of using search engines, assessing and selecting links, scanning and reading information resources, and using various backtracking mechanisms (e.g., history lists or back buttons on a browser).

It was noted that the overall information-seeking process in this context was highly *iterative* in nature and broadly organized into two major loops of activities: (1) *search* loop, and (2) *sense making* loop. The search loop involved things aimed at retrieving information resources, including formulating queries, submitting queries, evaluating search results for selection, editing queries, and filtering search results. The *sense making* loop involved the iterative development of a mental model (a conceptualization) of the workings of biological systems by evaluating selected information resources for use, and eventually consuming the useful information resources.

Post search activities

Once an article was found, the post-search activities included the following. (1) Re-representation of the information contents of that article: designers created summaries of the articles, which included the explanations of the working of the biological systems using the *Why*, *What*, and *How* structure. (2) Organization of the information resources: designers organized the articles by uploading them to their T-Square accounts for future reference. (3) Sharing the information resources: designers shared the found information resources and their summaries with the team members, usually through electronic means.

Challenges of online bio-inspiration seeking

Although the basic paradigm of information seeking in the context of online bio-inspiration seeking was not very different from other everyday information seeking, it seemed certainly more challenging. A number of difficulties were noted in the process of online information seeking for the purpose of finding biological sources of inspiration. These difficulties were encountered irrespective of the type of information environment and contributed greatly to the inefficiency of the information seeking process, causing designers to experience some degree of tedium and frustration. In particular, three difficulties surfaced prominently.

Low find frequency issue

Designers often go for long periods without finding a useful or relevant information resource in the information seeking process. In other words, the relative frequency of encountering useful information resources in this context was typically very low. This can be contrasted with our everyday online information seeking experiences where we frequently find useful information resources in response to our information needs and do so with relative ease. Anecdotal evidence for this issue can be found in comments such as these from designers:

"I really had a lot of trouble completing [the] assignment last night. They really need to come up with a better way for people for looking up information [based] on function. [I] Wasted so much time looking everywhere but found only one [article] which is just so-so"

"I am not about to give up on aquaporins because it took me a long time to find that damn thing"

A rough back-of-the-envelope calculation also suggests that designers spent approximately three person-hours of search time in order to find a single relevant article. Team FORO, consisting of 6 people, collected 39 articles over the 7 week period, where each designer reported spending an average of 2 to 3 hours per week on this task.

The find frequency depends on a combination of several factors, chief among them being the nature of the task in which the information seeking is embedded, the nature of the information environment, and the expertise of the information seeker. The low find frequency in this case suggests a lack of adequate affordance of Web-based online information environments for the task of bio-inspiration seeking for this group of designers.

Recognition error issue

Designers were prone to making errors in the judgment of the true utility of information resources that they encountered in the search process. It was noted that in almost all online environments, search queries brought back a ranked list of search results (a set of information resources). One important aspect of the search process was assessing and selecting promising information resources from this list for further consumption. But, this decision had to be made based on proximal cues (or information surrogates) – links

and pieces of text that are intended to represent the distal information resources. For instance, Figure 5.5 shows the proximal cues in Google Scholar.

Figure 5.5: Search results page of Google Scholar showing proximal cues

This decision-making about which information resources to pursue and which ones to reject (using proximal cues) was prone to error. In many instances, designers picked up on low-utility articles and spent a lot of time and effort trying to understand its contents, only to realize later that it was not actually very useful (false positives). In my own information seeking experiences, I tried to make an entry in my field notes every time such an occurrence registered on my conscience mind. There were approximately 53 such instances in my overall bio-inspiration seeking process over the seven-week period. When I raised this issue in one of the team meetings, my team members agreed that this was not uncommon. But when asked why this was the case, they tended to blame it on their insufficient knowledge of the biology domain. False positives lead to wasted time and effort (resource cost). False positives also have opportunity costs associated with them – by handling less profitable items one looses, in that time, the opportunity to go after more profitable items.

Conversely, one can image situations where designers might dismiss a resource (based on its surrogate) that they encounter during the search as having low utility even though in actuality it might have contained useful information about a potential biological source (false negatives). Unfortunately, even though such situations are highly probable, it is practically impossible to observe them in the field because the rejected items are not tracked and independently assessed for their true utility. However, in Chapter 10, we will show the occurrence of false positives in controlled settings. Although false negatives do not have resource cost associated with them, they represent lost opportunities.

The resource and opportunity cost of recognition errors, coupled with the fact that there is a tendency among designers to fixate on the biological sources that they find initially, can potentially lead to suboptimal choice of biological sources of inspiration.

The issue of understanding biology articles during search

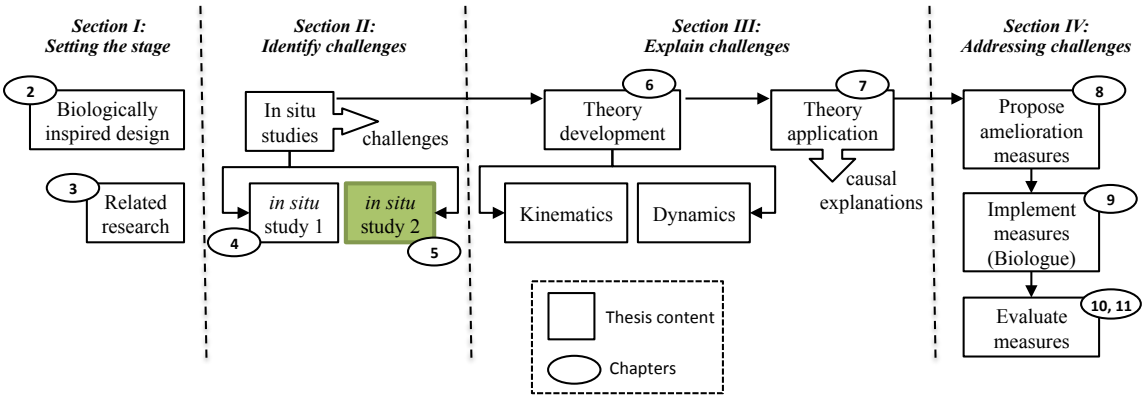
Design has its own distinct 'things to know, ways of knowing them (Cross 1982, p. 221).' However, designers often struggled with 'designerly' ways of coming to know biological systems during the information seeking process. Developing an understanding of the workings of unfamiliar biological systems from the information resources that one encounters during the information seeking process is an integral part of the online bio-inspiration seeking task. Furthermore, developing the right kind of understanding of these systems is also important – the kind that allows designers to "see" in what ways a biological system is similar or dissimilar to the technology that is being designed and how its mechanisms, strategies, principles, etc., can or cannot be transferred and adapted to solve the target design problem. This highlights the fact that learning and information seeking are inextricably intertwined in the context of bio-inspiration seeking task.

Coming to a reasonably good understanding of systems from online information resources presented one of the biggest challenges for designers (especially for non-biologists). That it was difficult to comprehend biology articles was one of the common

complaints that got expressed. This was in part due to the scholarly nature of the articles that were being sought and used. A majority of such articles are produced by experts and for experts in the biology community, whose focus is on communicating what, why, and how the researchers did their work, presenting key data (often in figures, tables, and charts), and discussing the results of their analysis. More importantly, the focus is *not* always on providing a step-by-step guide to how a biological system works. The level of abstraction at which the targeted biological systems are discussed in these articles are often too detail-oriented, obfuscating the "big picture" of their workings. This issue, combined with the technical and domain-specific nature of the vocabulary used, and the implicitness or omission of key concepts required for constructing understanding, hinder the sense making process and the construction of the kind of mental models of biological systems that are required for the design activity.

One consequence of this difficulty is that the process of sense making itself spawns off new information needs and contributes additional cycles of information. In other words, search cycles can lead to sense making issues, which in turn can lead to more search cycles. This property of information seeking adds to the complexity of online bio-inspiration seeking process and can sometimes vastly increase the cost structure of the process.

Chapter Summary



This chapter takes a closer look at the nature of online bio-inspiration seeking process and its associated challenges by means of a participatory research study. This research study analyzed the online bio-inspiration process in terms of the practices and products of one design team's effort to develop a novel bio-inspired desalination technology. In this study we noted that bio-inspiration seeking activity collaborative in nature, characterized by a shared goal and joint effort in accomplishing the goal. This process of bio-inspiration seeking consisted of three stages of activities. The pre-search was targeted towards a socially negotiated arrival of a common understanding of the target problem, establishment the information needs for the next stage and division of labor to fulfill those needs. The during-search stage consisted of individual team members engaging in online information seeking their own. This individual information seeking process was exploratory, iterative and consisted of searching and browsing and sense-making tasks that interacted opportunistically. The post-search stage was targeted towards re-representation and organization of found information resources. Both the information resources and their re-representations were shared among team members.

With respect to the challenges associated with online bio-inspiration seeking, three specific challenges were noted. First, designers experience *low rate of encountering relevant information resources* in online environments that they conventionally use (e.g., Google Scholar, Web of Science, Google, Wikipedia). Although designers are awash with biological information during online bio-inspirations seeking, they often go for long periods without encountering information resources that actually contain analogous biological systems. Second, designers experience *high rate of recognition errors*. Even though they encounter a large number of information resources during the search process, they often fail to recognize the actual relevancy of those resources. They make mistakes in identifying if an encountered information resource contains an analogous biological system or not, which lead to false positives and false negatives that can end up costing undue time and effort. Third, designers experience *significant difficulty in comprehending*

biology articles identified as being relevant. They struggle to develop conceptual understanding of biological systems discussed therein. In other words, designers find it challenging to build sufficiently deep mental models required to address their design challenges based on the contents of biology articles.

SECTION III: EXPLAINING THE CHALLENGES

Section Summary

The information presented in this section tries to address the research question:

- ***RQ2: What are the causes underlying the challenges of online bio-inspiration seeking?***

In order to answer RQ2, it is important to gain an in-depth understanding of the nature of the online bio-inspiration seeking phenomenon. In Chapter 6 of this section, I developed a theory of online bio-inspiration seeking which guides our understanding of online bio-inspiration seeking phenomenon. This theory provides two kinds of accounts of online bio-inspiration seeking.

First, the *kinematics* account provides a purely descriptive account of the phenomenon based on my analysis of the bio-inspiration seeking practice in the two *in situ* studies. Second, the *dynamics* account provides an explanatory account of the phenomenon in terms of its underlying causal processes or 'mechanisms.'

This dynamics account explains the online bio-inspiration seeking phenomenon in terms of an information-processing model that I developed called *Interactive Analogical Retrieval (IAR)*. The IAR model is a synthesis of two existing theoretical frameworks: one, a traditional cognitive model of analogical retrieval called Analogical Retrieval by Constraint Satisfaction (ARCS) (Thagard et. al., 1990), and two, a human-information interaction theory which explains people's online information-seeking behavior called Information Foraging Theory (Pirolli, 2007).

The IAR model can be used to reason forwards from deliberate changes in the information environment to its observable effects on the online bio-inspiration seeking process of designers, or backwards from observed bio-inspiration seeking effects to the factors in the information environment causing those effects. Reasoning backwards, the IAR model provides causal explanations for the three observed challenges associated

with the online bio-inspiration seeking process. These explanations are captured in Chapter 7.

Thesis 2: The causes for the identified challenges of online bio-inspiration seeking can be traced to three factors in the online environments that designers currently use. (2a) The issue of *low rate of encountering useful information resources* issue can be traced to the current keyword-based methods of indexing and accessing information resources in online information environments, which support access to information resources based on literal similarity (word-for-word matching) while ignoring semantic-, structural- and pragmatic-similarity – the three pressures governing the process of analogical retrieval. (2b) The issue of *high rate of recognition errors* issue can be traced to the nature of proximal cues that one customarily encounters in current online information environments – specifically, their lack of affordance for accurately perceiving the *information scent* (analogical similarity) of the resources they represent. (2c) The *conceptual understanding difficulty* issue can be attributed to the fact that existing biological information resources (especially scholarly articles) are usually created *by* and *for* (expert) biologists, which, for non-biologists, may not contain explanations at the right level of abstraction and may leave a lot of information implicit that constituting gaps in knowledge.

Once we have identified the causes, we can begin to address the question of how to address those challenges in the next section on ameliorating the challenges.

CHAPTER 6

KINEMATICS AND DYNAMICS OF ONLINE BIO-INSPIRATION SEEKING

This chapter develops a descriptive account (kinematics) and an explanatory account (dynamics) of online bio-inspiration seeking. Grounded in the two *in situ* studies discussed in the previous two chapters, these two accounts contribute towards a theoretical understanding of online bio-inspiration seeking. The accounts developed here will be subsequently used in later the chapters to understand the causes underlying the challenges associated with online bio-inspiration seeking and to develop hypotheses about measures that can be taken to mitigate those challenges.

Theory development: Guiding principles

Information systems research, especially the domain of HCI, has adopted a number of theoretical constructs in the past. These include theories, conceptual frameworks, and descriptive methods, as well as a variety of hybrid forms (Shapiro, 1994). Just a partial list includes: activity theory (Engeström, 1987; Kuutti, 1996; Nardi, 1996b; Bardram, 1997; Engeström et al., 1999), conversation analysis (Sacks et al., 1978; Frohlich and Luff, 1989; Sacks, 1992; Katzenberg and McDermott, 1994), coordination theory (Schmidt and Simone, 1996; Carstensen and Nielsen, 2000), distributed cognition theory (Rogers and Ellis, 1994; Hutchins, 1995b; Ackerman and Halverson, 1998), ethnomethodology (Garfinkel, 1967; Button, 1991; Rouncefield et al., 1994; Heath and Luff, 1996), grounded theory (Glaser and Strauss, 1967; Strauss and Corbin, 1998), situated action (Suchman, 1987; Schiff et al., 1997) and social/symbolic interactionism (Blumer, 1986; Fitzpatrick et al., 1996).

Most of these have been used to study and describe work settings and systems, but few explicitly approach the design of those systems. As Button and Dourish (1996) point out, closing the gap between critique and design is quite a challenge. Information systems design often turns to other methods to support the design process including: participatory design (Greenbaum and Kyng, 1991), and user centered design (Norman and Draper, 1986).

The value of developing a theoretical framework in the present context is consistent with the view expressed in Barthelmess and Anderson (2002).

"the value of any theory is not 'whether the theory or framework provides an objective representation of reality' (Bardram, 1998), but rather how well a theory can shape an object of study, highlighting relevant issues. In other words, a classification scheme is only useful to the point that it provides relevant insights about the objects it is applied to."

From this point of view, theories are like a pair of dark glasses. We put them on and the world is tinted. The change brings some objects into sharper contrast, while others fade into obscurity. Therefore, it is important to develop or adopt theories that bring those theoretical objects into focus that appropriate for the goals of the research.

From a pragmatic standpoint, the theoretical framework developed in the present context aims to provide four tangible benefits that are identified by Halversion (2002) as best practices for theory development. First, it should provide *descriptive* power. It should provide a conceptual framework that helps us make sense of and describe the world. This includes describing a work setting as well as critiquing an implementation of technology in that setting. Second, it should provide *rhetorical* power. Theory should help us talk about the world by naming important aspects of the conceptual structure and how it maps to the real world. This is both how we describe things to ourselves and how we communicate about it to others. Further, it should help us persuade others to accept our views. The third attribute is *inferential* power. We do want a theory to help us make

inferences. In some cases those inferences may be about phenomena that we have not yet understood sufficiently to know where or how to look. We may hope that inferences will lead to insights for design. Or we may want to predict the consequences of introducing change into a particular setting. An important fourth attribute has to do with *application* power: how we can apply the theory to the real world for essentially pragmatic reasons. Mostly this translates to our need to inform and guide system design. We need to describe and understand the world at the right level of analysis in order to bridge the gap from description to design.

By virtue of being grounded in *in situ* observations of the phenomenon, the kinematics account of online bio-inspiration seeking that is developed here is expected to provide descriptive and rhetorical power. On the other hand, by providing a causal explanatory account of the phenomenon, the dynamics of online bio-inspiration seeking that is developed here is expected to provide inferential and application power.

Kinematics of online bio-inspiration seeking

The kinematics account begins by acknowledging that the emergence of the Web and various search environments that index its contents has made information seeking in online information environments a daily activity for people. Whether at work, school, or play, people have come to expect instant access to information on any topic at any place and time. This expectation carries over to BID situations where designers cannot rely on their cognitive memories to obtain their biological sources of inspiration. In such situations, designers collaboratively engage in information seeking in online information environments to discover their biological sources of inspiration. However, the nature of online bio-inspiration seeking is unique and different from other everyday information seeking tasks that needs to be studied and described in its own right. The kinematics of online bio-inspiration seeking can be schematically depicted as in Figure 6.1.

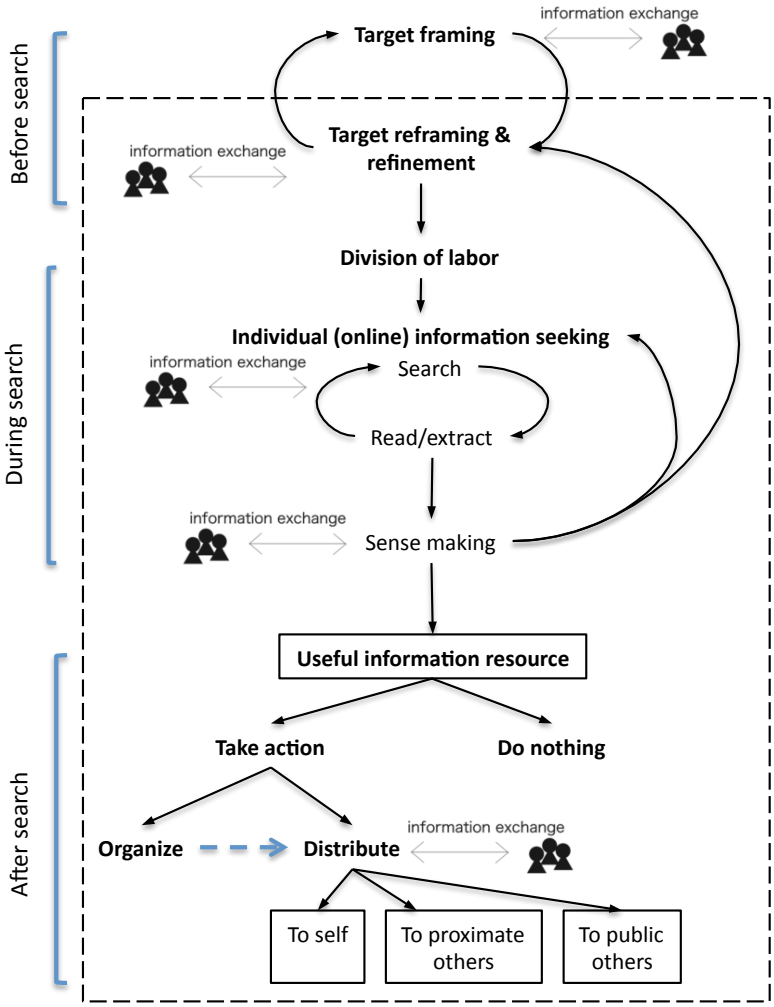


Figure 6.1: Kinematics of online bio-inspiration seeking

As depicted in the Figure 6.1, the process of online bio-inspiration seeking can be viewed as consisting of three stages of *before*, *during*, and *after* search stages. From what we have seen, online bio-inspiration seeking is a collaborative process (which does not mean it cannot be a solo activity). There are opportunities for collaboration in all the three stages of the process. The various points in the process where we have observed either direct or indirect collaboration are depicted in the process (in Figure 6.1) by this icon: 

Before search stage

Online bio-inspiration seeking is rooted in an information need - to find biological systems that can help solve a particular target design problem or better understand the problem. The before search stage consists of tasks that are geared towards *target framing* and *target reframing and refinement*, which refers to activities that help clearly establish the information need and refine those needs if required. The tasks in this stage are, strictly speaking, not part of the online bio-inspiration seeking process itself. However, the online bio-inspiration seeking process is situated in the past and present context of target framing and reframing, both influencing and being influenced by them. Therefore, it makes it hard to draw clear boundaries between what is part of the seeking process and what is not.

Target framing: The target design problem usually comes from external sources (e.g., clients, instructors) and other times emerge from the context the designer finds herself in. The events and interactions leading up to the target problem are included here. This is the precursor to the establishment of the information need. Traditional models of search and information seeking assume a well-established information need to begin with. Our observations show that the target framing in BID is mostly a social process involving negotiations among team members. Individual designers with similar interests are grouped together into teams. Each team is responsible for defining a problem that the team would address. Each individual has his or her own "pet" area that they wish to explore through their design problem (e.g., robotics, sensors). The team has to negotiate to come to a consensus about a concrete problem that takes into consideration each team member's inclinations. This negotiation process happens in design team meetings through face-to-face communication. Occasionally, an expert (instructor or mentor) mediates the target framing process, utilizing tools such as *functional decomposition* and *functional optimization* (refer to Chapter 5 for more details).

Target reframing and refinement: The target problem can be well understood or ill understood. Target reframing refers to a set of cognitive or social, intentional or non-

intentional events that shape and reshape the target problem. There are several cognitive models of problem solving that explain how this might occur given the background knowledge of the problem-solver (e.g. Clement 2008, Griffith, Nersessian & Goel 2000). But few of them consider the social channels that might lead to target reframing.

In the context of BID, target reframing is a necessary part of the information-seeking process because the target problem and the biological sources are situated in different domains. Designers initially define target problems in human terms, such as "protecting police" or "avoiding shark attacks." In order for designers to find solution analogues in biology, designers need to reframe their problem in more broadly applicable biological terms, often in the form of a question such as "How do biological systems accomplish [xyz] function?" Instructors term this reframing step as "biologizing" the problem. Target reframing is a social process. Designers in a team negotiate on what this biologized problem should be in team design meetings through face-to-face communication (refer to Chapter 5 for more details).

The biologized version of the problem can serve two purposes in the context of online bio-inspiration seeking. First, it can supply some of the keywords that designers use in the information seeking process. Second, the target framing and reframing activities can lead to the identification of multiple information needs which can then lead to the division of labor. This is again a negotiated social process (refer to Chapter 5 for more details).

During search stage

In the context of BID, it is very likely that designers will not have biological sources *a priori* encoded in their long-term memories and that they need to reach out to the external information environments, in particular the Web, to obtain them. Further, this search and retrieval process is not only distributed between the designer and the environment, but is also a social process.

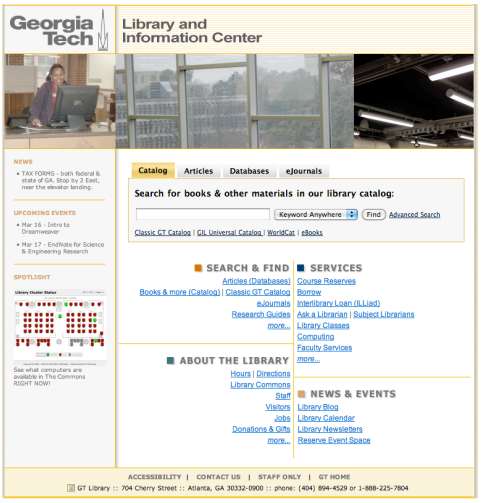
Using the heuristics and the biologized version of their target problem discussed in Chapter 4, designers seek information in the information ecology for articles in biology that discuss any biological system that might be relevant to meeting their information need for solving their target problem. The most commonly used Web-based environments to find the articles are Google, Google Scholar, Web of Science (recommended by the instructors), Wikipedia, and AskNature. This search process is carried out in parallel with each team member seeking information on his or her own first.

This individual information seeking process in this context is highly iterative in nature and broadly organized into two major loops of activities: (1) *search* loop, and (2) *sense making* loop. The search loop involves things aimed at retrieving information resources, including formulating queries, submitting queries, evaluating search results for selection, editing queries, and filtering search results. Often, submitted queries do not bring back useful information, prompting an edit action on the submitted query. There are two kinds of edits: *reformulation* and *refocus*. Reformulation involves rethinking and issuing fresh queries. Refocusing queries involves adjusting the focus of the existing query by adding, removing, or modifying one ore more terms. Whether reformulating or refocusing queries, the techniques of *generalization* and *specialization* are commonly employed. Making a search query more general usually involves removing or changing terms with the aim of retrieving a broader set of results, while making a query more specialized usually involves adding or changing terms with the aim of retrieving a narrower set of results.

Capturing a complex information need in a short query phrase that is likely to yield useful results is challenging. But once submitted queries typically bring back a ranked list of search results (a set of information resources). One important aspect of this is assessing and selecting promising information resources from this list for further consideration. Typically, this decision is made based on *proximal cues* (a.k.a information

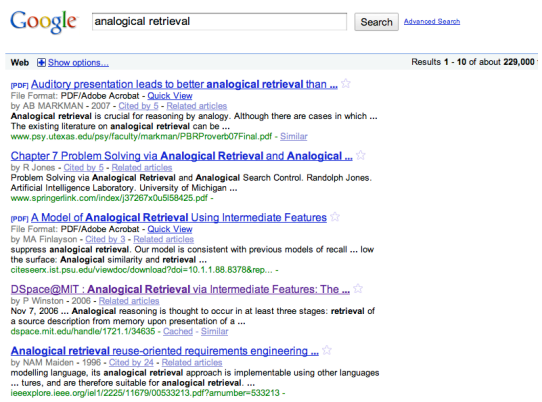
surrogates) – key frames for videos; sentences, titles or abstracts for documents; thumbnails for Web pages; and so on, which represent the distal information resources.

Figure 6.2 shows various types of proximal cues that one might come across in online information environments. They include, but are not limited to: underlined hyperlink text that one sees regularly on web pages and on Wikipedia pages (Figure 6.2a, Figure 6.2b), underlined hyperlink text that one sees on "Search Results" pages of search engines such as Google (Figure 6.2c), title and abstract and citation information associated with "Search Results" pages of bibliographic search engines such as Google Scholar (Figure 6.2d) and of digital libraries such as ACM Digital Library (Figure 6.2e), citation information within articles (Figure 6.2f).

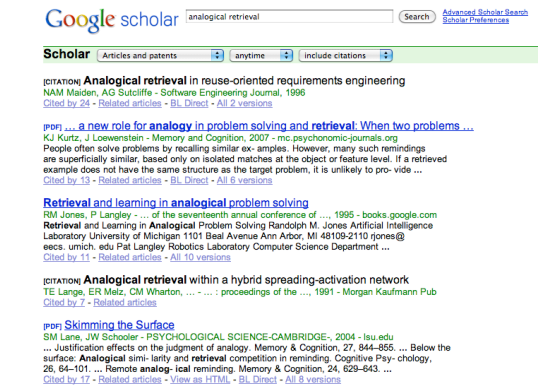


(a) A regular website; text on blue underlined links are the cues

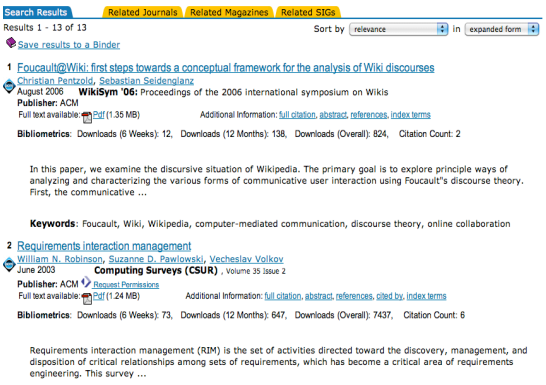
(b) A Wikipedia webpage; the text on blue links are the cues



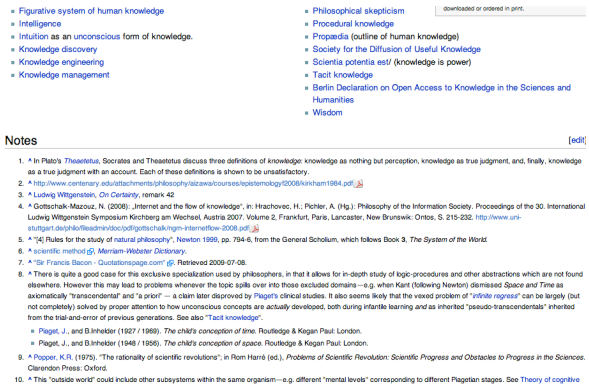
(c) Search results page of Google; text on the blue hyperlinks plus the snippets of text following each link are the cues.



(d) Search results page of Google Scholar; titles of articles, snippets of abstracts following the title, and the citation information are the cues.



(e) Search results page of ACM Digital Library; titles, snippets of abstracts, and the citation information are the cues.



(f) Citation information found within an article are the cues that can lead to other full articles (patches)

Figure 6.2: Proximal cues in different information environments

The *sense-making* loop involved the iterative development of a mental model (a conceptualization) of the workings of biological systems by evaluating selected information resources for use, and eventually consuming the evaluated information resources. This loop of activities consists of evaluating an information resource for utility, evaluating for personal fit, and its eventual use. The complexity of the evaluation strategies emerged from recognizing the three evaluation groupings (evaluation for

selection, *utility*, and *personal fit*) as interaction strategies, each with its own purpose and characteristics as summarized in Table 6.1.

Table 6.1: Information interaction strategies in kinematics of online bio-inspiration seeking

	Categories of information interaction strategies				
	Search	**Evaluation for selection**	**Evaluation for use**		**Use**
			Utility	**Personal fit**	
Purpose	Finding/locating information resources	Selecting resources to look at	Deciding what resources to read (examine in depth)	Deciding if the content of the resources can potentially satisfy information needs	Organizing and incorporating into a body of knowledge
Characteristics	Interaction with query interface, search engine, tools, etc.	Interaction with the result list, proximal cues, etc.	Interaction with the information resource, predominantly the initial parts	A more exhaustive interaction with the information resource (and other resources for reference). May require multiple passes	Interaction with the information resource, creation of external representations, note-taking, organization and integration with other resources and representations

Evaluation for *selection*, mentioned earlier in the search loop, involved interactions with proximal cues, while evaluation for *use* involved interactions with actual information resources themselves. Evaluation for *use* can be further categorized

into evaluation for *utility* and evaluation for *personal fit*. The transition from *utility* to *personal fit* was roughly distinguishable by: (1) a trend towards more intense, iterative interactions, e.g. in-depth source reading and multi-passes of sources, leading to identification of new, and clarification of already identified, conceptual extractions; (2) the emergence of examples of many 'on-source' external representations, e.g. highlighter marks and margin notes on the resource (PDF document) to emphasize for further attention etc., as well as 'off-source' external representations, e.g. hand-written/electronic descriptive notes for providing quick recall of main points etc.

The boundary between 'Evaluation' and 'Use' are somewhat blurry. But the *use* category signaled a change of task focus: rather than searching for sources, or searching the located sources for conceptual and physical extractions, the focus was on *mental model construction*. The term "mental model" has been used in many contexts for many different purposes. But in the present context it refers to a designer's construction of the understanding of the workings of a biological system by engaging with an information resource like a biology article. Other characteristics of this *use* category include: (1) the move to multi source interactions e.g. comparisons across many documents/articles; (2) the emergence of more complex, analytic external representations in the process of consuming an information resource, e.g. Why-What-How summaries of articles, multi-dimensional tables, graphs, diagrams, equations, etc.; (3) the integration of new-found information into existing body of knowledge of the team.

Although team members pursue information seeking as on their own, there may be many opportunities for social interactions among team members in order to coordinate their search efforts. One kind of interactions are carried out to implement division of labor within the group in order to increase the coverage of the vast information resource landscape and increase their chances of finding fruitful information patches to forage in (e.g., "I'll search Wikipedia, you search Web of science" or "I'll search the process of thermal desalination, you search membrane-based processes"). The other kind of

interactions is related to suggesting and sharing keywords to use for search queries. The third kind of social interaction is related to sense making of a potential resource. In the course of seeking information resources, designers have to peruse several potential articles and make decision about their utility. Often, in BID, designers do not have the requisite background knowledge in biology to make sense of an article. In such situations, designers seek clarifications from other team members. These interactions are ad hoc and happens in variety of ways, including classroom talk, email, telephonic conversations, and instant messaging. These opportunities for interactions are depicted in Figure 6.1 using the collaboration icon.

After search stage

Organization: Following the active search phase, an information resource (e.g., biology article) is often obtained, which may be "acted" on through organization and/or distribution. The organization of information includes: saving or augmenting materials, bookmarking a webpage, or creating a new document or presentation. This process is referred to this process as schematizing (Pirolli, 2007), where raw evidence is organized and ''represented in some schematic way." Some of the created artifacts based on search products that serve to organize, save, or synthesize important information include: printing results, bookmarking, creating summaries, etc. Our observations show that, many designers organize their found information resources in some fashion. The most common way was assignment-wise organization (e.g., problem description assignment, project 1 assignment). This was done in T-Square, the classroom resource management tool that they were required to use.

Distribution: Such organizational acts additionally served to distribute the search products to others. In our observations, many summaries and presentations were created with the intention of sharing their fruits of the information seeking process with colleagues. Pirolli (2007) observes that the end products of a search may be delivered to

an audience as a "presentation or publication of a case." It is also noted that users serve as information filters for others through their organizational, and consequent distributional, acts of bookmarking, tagging, or annotating items. Our observations show that most of the organizational acts served as a means for distributing the found information resources among team members. This led to the creation of a shared pool of information resources (in T-Square) that designers could tap into as and when needed. These activities suggest that social interactions are important even after the primary search act.

From kinematics to dynamics: Zooming in

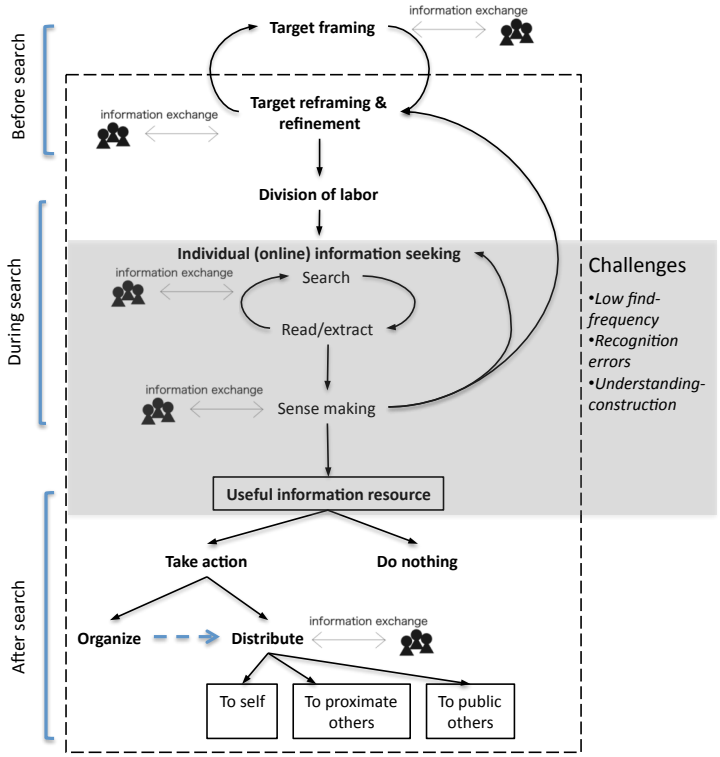


Figure 6.3: Observed challenges and their relation to the kinematics account

As we move from descriptive to explanatory account, we zoom into that part of the process that we are interested in explaining, which is indicated by the gray area in Figure 6.3. The reason we scope the explanation around this part is because we are interested in finding the causes and proposing hypotheses for the three identified

challenges: low frequency of encountering relevant information resources, recognition errors, and understanding-construction challenges (refer to Chapter 5 for more details).

Description versus explanation

The discussion of kinematics and dynamics in the present context hinges on the distinction between description and explanation, which need to be demarcated more clearly. The problem is a difficult one because, among other things, virtually any description (e.g., "it is raining") can serve as an explanation or be part of an explanation ("Why do you take your umbrella?" – "Because *it is raining*"). How, when, and why a description qualifies as an explanation, and in particular a good explanation, remains a matter of philosophical debate. Tonneau (2008) claims that:

"On virtually any view of scientific explanation, one explains a phenomenon (event, happening, state of affairs) by describing another. Of course, one does not explain a phenomenon (B) by describing it; rather, what one must do in order to explain B is to describe a phenomenon A distinct from B. Why is the temperature lower today in northern France (B)? Because of a mass of cold air coming from Russia (A). Why do objects accelerate in an inertial frame (B)? Because they are subject to forces (A)."

In the present context, the phenomena under consideration is online bio-inspiration seeking. The kinematics account provides a description of the phenomenon by synthesizing the observations and findings of the two *in situ* studies discussed in the previous two chapters. This refers to B, the state of affairs, in the above quotation. On the other hand, the dynamics account provides an explanation of the observed phenomenon by describing another process, *interactive analogical retrieval*, which represents A in the above quotation. Therefore, we have:

- The phenomenon: online bio-inspiration seeking
- The kinematics of the phenomenon (B): a description of online bio-inspiration seeking process based on empirical observations

- The dynamics of the phenomenon (A): a description of interactive analogical retrieval process
- B is explained in terms of A: our observations of online bio-inspiration seeking are explained in terms of interactive analogical retrieval theory.

However, one will find that there is some degree of overlap between the kinematics and dynamics accounts of online bio-inspiration seeking. The requirement that A and B be distinct does not prohibit B from overlapping with, or even being a proper part of, A. For instance, when I explain the polar boars' fur color in terms of natural selection, I attribute fur color (B) to a complex state of affairs (A), *natural selection of fur color*, that has B as one of its parts. Yet, clearly the two states of affairs, A and B, are distinct: the first (natural selection for fur color) includes reproductive success as a component, whereas the second (fur color) does not.

Dynamics of online bio-inspiration seeking

While the kinematics of a phenomenon provides a descriptive account of the phenomenon from the vantage point of an external observational investigation, the dynamics provides an explanatory account of the phenomenon in terms of the causal processes or 'mechanisms' underlying the phenomenon. The dynamics account developed here explains the online bio-inspiration seeking phenomenon in terms of the process of *Interactive Analogical Retrieval (IAR)*.

Biologically inspired design is a kind of analogical design. The task of seeking bio-inspiration involves obtaining biological analogues to target design problems. Therefore, the task of seeking bio-inspiration is a kind of analogical retrieval. However, online bio-inspiration seeking differs from conventional accounts of analogical retrieval in cognitive science. Analogical retrieval has traditionally been portrayed as an internal cognitive process of an individual mind, occurring over the long-term memory, with little room for the external environment in its description. In contrast, online bio-inspiration

seeking involves the individual, the external information environment, and the interaction between the two. Therefore, a process account of online bio-inspiration seeking should not only incorporate elements from traditional analogical retrieval theory, but also include elements of human-information interaction theory.

Interactive analogical retrieval account developed here to explain the dynamics of bio-inspiration seeking requires two building blocks: one, an account of human information seeking in online information environments, and two, an account of analogical retrieval. In the following sections, these two building blocks are discussed.

Information foraging theory: The first building block

Information foraging theory (Pirolli, 2007) is a theory that explains how people interactively seek information in their environment, particularly online information environments. The emergence of global information ecology has created enormous pressures for people who seek useful information. This theory provides insights into how people adapt to these pressures. Information Foraging Theory is grounded in computational theories of human cognition and optimal foraging theories in biology. Information foraging theory arose in the 1990s, coinciding with the explosion in the amount of information that became available to the people, and with the development of new tools for accessing and interacting with information online. The scientific foundations of this theory lay in the *adaptionist* paradigm where users are viewed as complex adaptive agents who shape their actions and strategies to be more efficient and functional in the information ecology. Roughly speaking, the adaptionist paradigm involves asking what environmental problems are being solved and why cognitive and perceptual systems are well adapted to solving those problems. Information foraging theory rests upon certain axioms enumerated below.

Explanations of information seeking behavior based on food foraging mechanisms in biology: Natural selection favored organisms - including our human

ancestors - that had better mechanisms for extracting energy from the environment and translating that energy into reproductive success. Our ancestors evolved perceptual and cognitive mechanisms and strategies that were well adapted to the task of exploring the environment and finding and gathering food. Information foraging theory assumes that modern-day information foragers use perceptual and cognitive mechanisms that carry over from the evolution of food foraging adaptations.

Relevance of optimal foraging theory and models from biology: If information foraging is like food foraging, then models of optimal foraging developed in the study of animal behavior and anthropology should be relevant. A typical optimal foraging model characterizes an agent's interaction with the environment as an optimal solution to the tradeoff of costs of finding, choosing, and handling food against the energetic benefit gained from the food. Information foraging models, too, include optimality analysis of different information seeking strategies and technologies as a way of understanding the design rationale.

The economics of attention and the cost structure of information: Herbert Simon remarked that the wealth of information creates poverty of attention and an increased need to efficiently allocate that attention. In an information-rich world, the real problem is not so much how to collect more information, but, rather, to increase the amount of relevant information encountered by the information seeker as a function of the amount of time that the user invests in interacting with the information environment. The more the environment affords a seeker to attend to relevant information per unit time the better the amplification of cognition that is achieved. The structure of the information environment (both physical and online) determines the time costs, resource costs, and the opportunity costs associated with exploring and exploiting information. People have limited attention and must deploy that attention in a way that is adaptive to the task of finding valuable information. People prefer information seeking strategies that yield more useful information per unit time, and they tend to restructure their environments to

optimize the rate of gain. People also prefer, and consequently select, technological designs that improve returns on information foraging.

The inherent patchy nature of the (online) information environment: Information seekers usually surf the online environment (Web) seeking content related to some topic of interest, and the online environment is conceptually and structurally organized into topical localities. The local groups of information are generically referred to as *information patches* - the idea being that it is easier to navigate and process information that resides within the same patch than to navigate and process information across patches. In other words, the term "patch" suggests a locality in which within-patch distances are smaller than between-patch distances. Examples of patches include individual web pages in the hyper-linked World Wide Web, individual documents in online document repositories, individual articles in digital libraries, individual entries in online encyclopedic resources such as Wikipedia, etc. Empirical studies of the structure of the online environments confirm a patchy structure (see Pirolli, 2007).

A high-level process account of information foraging

Information foraging theory is a framework that explains how people seek information in external information environments. In the domain of information seeking, the predator is the person in need of information and the prey is the information itself. Just as predators seek prey for the least possible effort, information foragers seek relevant information for the least possible interaction cost.

Figure 6.4 depicts the information seeking process (in analogy to food foraging) within the framework of information foraging theory. According to this process, upon probing information environment with some information goal in mind, a forager encounters numerous information patches (e.g., Web pages, online articles, etc.) that compete for forager's attention. These patches may or may not contain information relevant to the forager's goals. Forager expends some amount of time and effort

navigating from one patch to another until one that can be exploited is found. This search process is referred to as *between-patch* foraging. Between-patch foraging costs can vary from little or no cost (if one knows exactly what information is sought and where to find it) to costs that are frustratingly high.

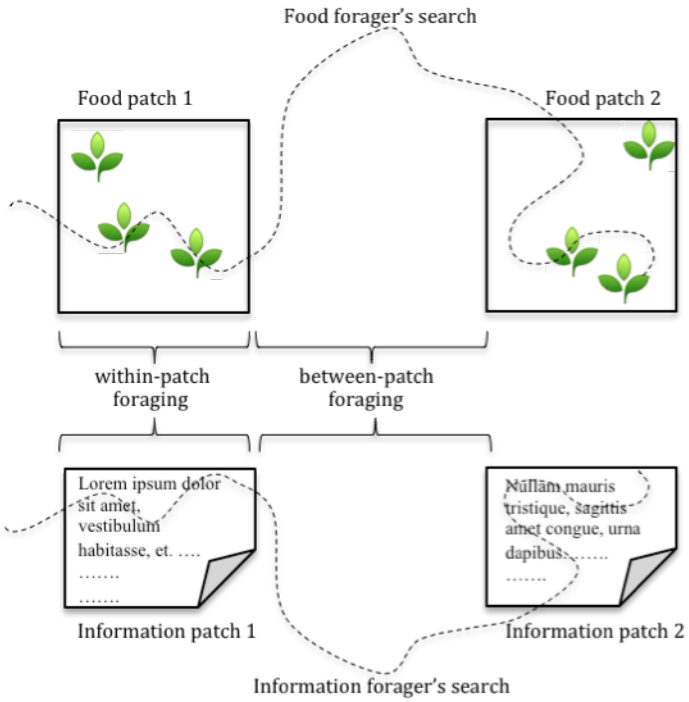


Figure 6.4: Information foraging process (in relation to food foraging)

Once a patch is selected, the forager starts consuming the information in that patch and constantly faces the decision of continuing to forage in the current patch or leave it to seek another patch. This is referred to as *within-patch* foraging. This decision to continue or leave is based on how much useful information the patch is currently yielding, weighed against the possibility of finding better patches should the forager choose to leave. Patches usually yield diminishing information returns over consumption time.

Between-patch foraging using information scent

The structure of the online information environments have evolved to exhibit certain regularities in the distribution of information resources and the navigation mechanisms that lead to those resources. One such regularity is that when foragers encounter patches in the online information environment, they cannot perceive the contents of those patches all at once. Rather they perceive snippets of information representative of the distal information patches. These snippets of information are referred to as *proximal cues* or *scent cues* - cues that users can perceive in their local information environment to judge the utility of distal information patches and can choose to either navigate towards or away from those patches. Proximal cues are intended to represent tersely the content that a forager will encounter by choosing a particular patch.

The perception of proximal cues associated with information patches is referred to as *information scent* of a patch. Information scent is also a measure of the perceived relevance of an information patch based on the cues. If proximal cues are perceived to have high information scent, a forager will assess that the patch associated with that scent is likely to lead to information relevant to forager's goals and vice versa.

One crucial aspect of the above process relates to how a forager computes the information scent of a patch based on the perceived proximal cues, given forager's information goals. This step is crucial because a forager's decision to attend to a patch or not rests on strength of the information scent. Information foraging theory presents an *information scent model* that explains how this scent is computed. This model is based on the assumption that the goal of the information forager is to use the information scent obtained from proximal cues (e.g., a Web link) to predict the utility of distal information patches (e.g., the Web page associated with a Web link), and to choose to navigate those links having the maximum expected utility. This model is grounded in (1) a Bayesian analysis of the expected relevance of a distal source of content conditional on the available information scent cues, (2) a mapping of this Bayesian model of information scent onto a mathematical formulation of spreading activation mechanism, and (3) a

model of rational choice that uses spreading activation to evaluate the utility of alternative choices of navigating to different patches under consideration.

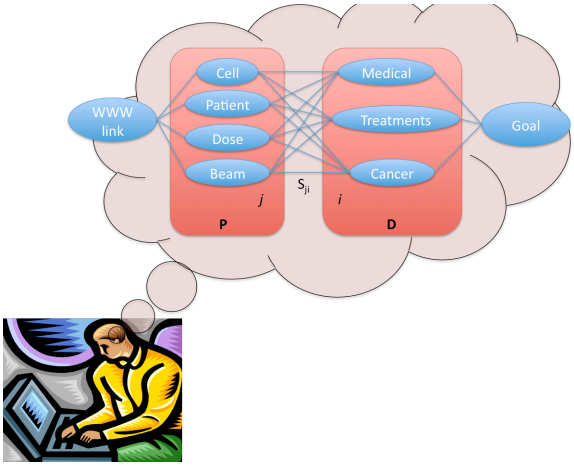


Figure 6.5: A schematic example of the information scent assessment

The spreading activation model of information scent assumes that the forager's cognitive system represents information scent cues and information goals in cognitive structures called chunks. Figure 6.5 presents a schematic example of the information scent assessment subtask facing a Web user. Figure 6.5 assumes that a user has the goal of finding information about "medical treatments for cancer," and encounters a Web link labeled with the text that includes "cell", "patient", "dose", and "beam". The user's cognitive task is to predict the likelihood that this distal source of content (a Web page) contains the desired information based on the proximal information scent cues available just in the Web link labels. Each node in the thought bubble in Figure 6.5 represents a cognitive chunk. Chunks representing information scent cues are presented on the left side of the thought bubble, chunks representing the user's information need are presented on the right side. Also represented by lines in the figure are associations among the chunks. The associations among chunks come from past experience. The strength of associations reflects the degree to which proximal information scent cues predict the

occurrence of unobserved features. The strength of association between a chunk *i* and chunk *j* is computed as:

$$S_{ji} = \log(\frac{\Pr(i \mid j)}{\Pr(i)}), \qquad \text{(EQ.6.1)}$$

where $\Pr(i \mid j)$ is the probability (based on past experience) that chunk *i* has occurred when chunk *j* has occurred in the environment, and $\Pr(i)$ is the base rate probability of chunk *i* occurring in the environment.

It is assumed that when a forager focuses attention on information scent cues, it activates corresponding cognitive chunks. Activation spreads from those attended chunks along associations to related chunks. For instance, activation would flow from the chunks on the left of thought bubble through associations to chunks on the right of the bubble. The amount of activation accumulating on the representation of a user's information goal provides an indicator of the likelihood that a distal source of information has desirable features based on the information scent cues immediately available to the user. For each chunk *i* involved in the user's goal, the accumulated activation received from all associated information scents chunks *j* is:

$$A_i = \sum_j W_j S_{ji} \qquad \text{(EQ.6.2)}$$

where W_j represents the amount of attention devoted to chunk *j*. The total amount of activation received by all goal chunks *i* is:

$$V = \sum_i A_i \qquad \text{(EQ.6.3)}$$

It is assumed that the utility of choosing a particular link is just the sum of activation it receives (equation EQ.6.3) plus some random noise. Based on this assumption it can derived that the probability that a user will click on a link *L*, having a summed activation V_L, from a set of competing links, *C*, given an information goal, *G*, to be:

$$\Pr(L \mid G,C) = \frac{e^{\mu V_L}}{\sum_{k \in C} e^{\mu V_k}} \qquad \text{(EQ.6.4)}$$

Within-patch foraging

Once the forager picks up scent of a potentially useful information patch, the forager goes to that patch and starts consuming information in that patch. One decision facing a forager in the within-patch foraging mode is whether to continue consuming the current patch or leave to search for another patch. This decision-making has to be adaptive to two pressures.

First, information patches typically yield diminishing returns with respect to the amount of information gained versus the time spent consuming a patch. That is, if a forager is in a patch not encountered before, the forager would at first find a relatively rapid increase in the amount of information gained, followed by fewer gains as the consumption time progresses. This is analogous to the property of food patches – as an animal forages within a food patch, the amount of food diminishes or depletes. For instance, a bird might deplete the berries on a bush as it eats them.

Second, the amount of time spent within a patch has to be weighed against the *opportunity cost* of finding a better patch. In the bird foraging analogy, for example, as the yield of a food patch decreases (as a result of bird's consumption of berries in the bush), there will be a point at which the expected future gains from foraging within the current bush diminishes to the point that they are less than the expected gains that could be made by leaving the current bush and searching for a new one.

To capture this adaptive decision-making process, Information Foraging Theory assumes that foragers employ learning mechanisms gained through experience to develop an assessment of the potential *information yield* of different kinds of information patches. For instance, foragers might learn over time that, given a current state of information need, Wikipedia web pages yield *x* amount of information for them, magazine articles

yield *y* amount of information, an article in Journal Science yields *z* amount of information, and so on. The balance between within-patch foraging and opportunity cost is captured using the *potential function*, $h(x)$:

$$h(x) = U(x) - C(t), \qquad \text{(EQ.6.5)}$$

where x is the forager's current experiential state (including the information goal), $U(x)$ is the utility of continued foraging in the current information patch, and $c(t)$ is the opportunity cost of foraging for the *t* amount of time that is expected to be spent in the current information patch. So long as the potential of the current information patch is positive (the utility of continuing is greater than the opportunity cost) then the forager will continue foraging in the current patch.

Charnov's Marginal Value Theorem

This theorem was developed to deal with the analysis of within-patch time allocation problem for food patches that yield diminishing returns curves, such as one depicted in Figure 6.6(a). Specifically, it tries to determine the optimal allocation time for within-patch foraging - how long should a forager spend consuming a patch to be optimal, considering the two pressures mentioned above. Let:

t_B = average between-patch foraging time,

t_W = within-patch for

$g(t_W)$ = information gain curve of a patch, which is a function of the within-patch foraging time,

$\lambda = 1/t_B$ = prevalence of patches in the environment captured by the inverse of the mean between-patch search time, t_B,

$R(t_W)$ = average rate of gain for the environment, such that

$$R(t_W) = \frac{\lambda g(t_W)}{1 + \lambda t_W} \qquad \text{(EQ.6.6)}$$

Then:

$g'(t^*) = R(t^*) = \frac{\lambda g(t^*)}{1 + \lambda t^*}$ where

t^* = optimal within-patch foraging time,

g' = the marginal value (the derivative) of the gain function g.

This theorem implies that a forager should remain in a patch as long as the slope of $g(t_W)$ (i.e., the marginal value of $g(t_W)$) is greater than the average rate of gain for the environment $R(t_W)$.

Figure 6.6(a) shows graphical representation of this theorem that appears in many discussions of optimal foraging theory. To determine the optimal rate of gain, *R**, one draws a tangent to the gain function ($g(t_W)$) and passing through t_B to the left of the origin. The slope of the tangent will be the optimal rate of gain, *R*. The point of tangency will also provide the optimal allocation to within-patch foraging time, *t**. The point of tangency is the point at which the slope (marginal value) of $g(t_W)$ is equal to the slope of the tangent line, which is the average rate of gain *R*.

From this theorem we can infer two features of foraging. First, reducing the between-patch foraging time (Figure 6.6(b)) not only increases the rate of information gain but also decreases the optimal allocation time to within-patch foraging. Second, increasing the slope of the gain curve (Figure 6.6(c)) not only (1) increases the rate of information gain (or information yield), and (2) decreases the optimal allocation time to within-patch foraging, but also (3) increases the amount of information gained.

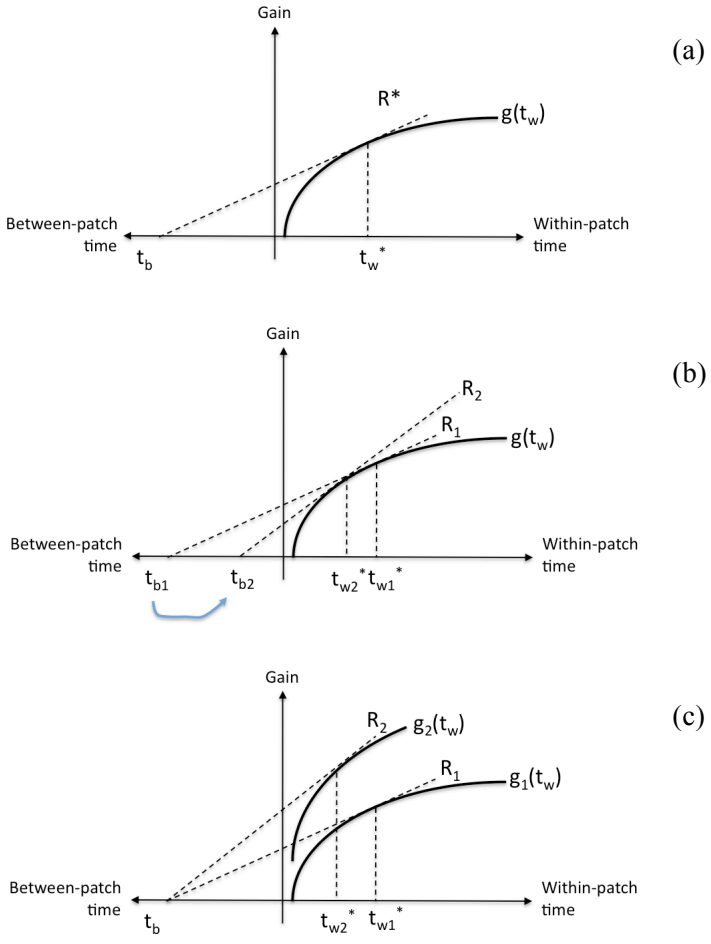


Figure 6.6: A graphical representation of Charnov's Marginal Value Theorem. (a) The rate-maximizing time to spend in patch, t*, occurs when the slope of the within-patch gain function (g) is equal to the average rate of gain. (b) The average rate of gain increases with decreases in between-patch time costs. (c) Under certain conditions, improvement in the gain function also increase the average rate of gain.

Analogical retrieval theory: The second building block

Analogy is the process of inferring something about one concept, the *target concept*, based on its similarities to another concept, the *source (or base) concept*. Analogy involves identifying conditions that hold for both the source and the target, and then inferring some additional condition that holds in the source that might also hold in the target. According to one school of thought, we can decompose the process of analogy into one or more of the following sub-processes (Gentner, 1999):

1. Retrieval or selection of the source concept

2. Mapping between the source and the target concepts
3. Transfer of information in the source for appropriate use with the target
4. Subsequent learning

An alternative school of thought proposes a comparable set of sub-processes (Kolodner 1993; Leake, 1996):

1. Situation assessment, which elaborates the problem situation and characterizes it in a form that is compatible with indexes used in the memory
2. Retrieval or selection of a source
3. Mapping process, which establishes correspondences between the target and source
4. Similarity assessment, which identify relevant similarities and differences
5. Adaptation process, that fit the retrieved source to the target situation
6. Evaluation process, which identifies the problems in the result and guides the incremental adaptation or further justification

Irrespective of the school of thought, retrieval is considered one of the key processes of analogy, including analogical design tasks such as biologically inspired design. Given a target problem, concept, or situation the process of analogical retrieval has been traditionally defined as the process of accessing and attending to a prior similar concept (or situation). An assumption that is implicit in this definition is that the process of analogical retrieval is an in-memory process involving working and long-term memories.

There are several theories that have been used to explain the process of retrieval emerging from both schools of thought. I will briefly discuss them here as psychological theories and AI theories. The distinction between the two categories are meant to capture the originating discipline as a convenient label, but not to suggest that psychological theories are devoid of computational elements and vice versa.

Psychological accounts of analogical retrieval have resulted in chiefly two important information-processing models of retrieval. One, MAC/FAC (Forbus & Gentner, 1995) is a model that was developed to address the following properties of analogy: (1) structural commonalities are weighed more heavily than surface commonalities in soundness or similarity judgments; (2) superficial similarity is more important in retrieval from long-term memory than is structural similarity; and yet (3) purely structural (analogical) remindings are sometimes experienced. MAC/FAC (for "many are called but few are chosen") consists of two stages. The first stage (MAC) uses a computationally cheap, non-structural match to filter candidates from a pool of memory items. Structured representations are redundantly encoded as content vectors, whose dot product yields an estimate of how well the corresponding structural representations will match. The second stage (FAC) uses structure-mapping engine (SME) (Gentner 1983; Falkenhainer, Forbus & Gentner, 1989) to compute a true structural match between the probe and output from the first stage.

The second psychological account, ARCS (Thagard et. al., 1990) (for “analogical retrieval by constraint satisfaction”), claims that structural similarity, semantic similarity, and pragmatic importance determine a set of constraints to be simultaneously satisfied during retrieval. It proposes an architecture in which the source retrieval is an emergent result of constrained, parallel activation of states of in a connectionist structure. This model is supplied with representations of the target and source and proceeds to build a localist constraint-satisfaction network in which hypothesis nodes correspond to possible hypotheses pairing the elements of the source with those of the target. Excitatory and inhibitory links between these nodes implement the constraints. In this way, contradictory hypothesis nodes compete with one another and (usually) do not become simultaneously active, whereas consistent nodes tend to mutually support each other. When the network settles, it provides a parallel evaluation of all possible mappings and finds the best one, represented by the set of most active hypothesis nodes. ARCS is coupled with a mapping

engine ACME (analogical mapping by constraint satisfaction) (Holyoak & Thagard, 1989) to produce the observed effect in which mapping is dominated by structural similarity and retrieval is dominated by superficial similarity.

The AI theories of retrieval are related to the paradigm of Case-based reasoning (CBR). The *case-based reasoning* approach is founded on the view that much human reasoning is based on reminding and use of specific episodes in memory (Kolodner, 1993). Influenced by Schanks' (1982) theory of dynamic memory organization and access, this approach focuses on how memory can be organized and indexed such that relevant episodes are retrieved when needed. Prominent examples of this approach include Kolodner's (1983) model of retrieval (implemented in CYRUS) which uses discrimination-net approach to selecting cases from a memory organized in terms of E-MOPS (episodic memory organization packets); Hammond's (1986) model (implemented in CHEF) which uses a memory organization consisting of plans, indexed by successes, failures, modifications, and repair strategies with respect to the goal accomplishment; Carbonell's (1986) model (implemented in ARIES) which uses a memory model consisting of cases containing traces (or history) of problem-solving episodes indexed by the major structural and control decisions; Winston's (1980) model of analogical retrieval which operates by moving down an annotated A-KIND-OF hierarchy of frames representing situations, with the slots in the frames "voting" concerning their relevance to the probe situation.

In the domain of case-based design, proposals of memories organized around design knowledge consisting of generalized domain knowledge like function models, indexed and retrieved by function, behavior and structure features are quite common (e.g. Goel, 1991; Bhatta & Goel, 1997; Gero, 1990; Qian & Gero, 1996). Also common are proposals of design memories organized around specific design cases, indexed and retrieved by features like design requirements, issues, pitfalls, stakeholders etc. (e.g.,

Domeshek & Kolodner, 1996). Design memories that combine both generalized design knowledge and specific cases have also been proposed (e.g., Maher & Zhang, 1993; Maher & Gomez de Silva Garza, 1996).

Which account to choose?

Any analogical retrieval account that serves as a building block here must explain the observations noted in the case studies (described in Chapters 5 and 6):

The fundamental unit of the analogical retrieval system is no longer confined to the mind of the individual analogist. The retrieval system comprises of the designer, the external information environment, and the interactions between the two. This requires a shift in the unit of analysis to include the environment as a first class object in the explanation.

This retrieval system often retrieves biological source analogues that are *superficially similar* to the target problem. For example, designers looking to harvest water in desert environments from fog might retrieve and attend to biological organisms like water-holding frogs (adapted to survive in deserts by forming an external water-proof cocoon to reduce water loss) because they have certain common surface features like water, extreme heat, desert, conservation, survival, etc., while lacking any deeper commonality in terms of mechanisms and principles.

This retrieval system also retrieves biological source analogues that apparently lack any superficial similarity to the target problem but share a *deeper* similarity in terms of shared abstract relationships like functions, mechanisms and principles. For instance, the same designers trying to design a fog-harvesting device in desert environments might retrieve and attend to models of airflow, humidity, and temperature regulation in tropical plants, although these two systems might not have anything in common in terms of their superficial characteristics; the similarity between the two might lie at the level of physical processes and applicable domain principles like the laws of fluid and thermo dynamics.

Finally, this retrieval system retrieves biological source analogues that are more or less *feasible* in terms of replicating, adapting, and scaling the sources' materials and mechanisms in order to generate the desired design solution. In some cases, it might be relatively straightforward to implement a solution based on the biological source. In other cases, it might be infeasible to implement a solution because the biological materials and mechanisms are either difficult to replicate in the engineering world and/or are infeasible to scale to the human scale.

Superficial similarity, deep similarity, and feasibility appear to act as constraints on the retrieval process. The following Table 6.2 samples the retrieved biological systems from the team FORO case study (see the case report of Chapter 5 for more details) and compares them to team FORO's target problem (to design an energy-efficient desalination technique) w.r.t. these three constraints.

Table 6.2: Biological systems retrieved by team FORO and their nature of similarity

Biological source	Superficial similarity	Deep similarity	Feasibility
Salmon gills	Yes	Yes	No
Penguin salt glands	Yes	Yes	No
Camel nasal turbinates	No	Yes	No
Kidney	No	Yes	No
Mangroves roots	Yes	Yes	Could not establish
Gobius Niger esophagus	Yes	No	No
Eel esophagus	Yes	No	No
Aquaporins	No	Yes	No
Small intestine	No	Yes	Yes

As the Table 6.2 shows, in some cases superficial similarity trumps deep similarity and vice versa during retrieval. But in a majority of the cases, both are satisfied but are trumped by the pragmatics of implementation. In the small intestine case (gray shaded row in the table), which was the only one judged as pragmatically relevant, only the deep similarity mattered, but the lack of superficial similarity did not matter.

We can apply Thagard et. al.'s (1990) ARCS model of analogical retrieval to provide one explanation for the aforementioned observations 2, 3, and 4. There is an obvious correspondence between semantic, structural, and pragmatic similarity in this model and the superficial similarity, deep similarity, and feasibility aspects of observed retrieval results. For this reason, this model provides a good starting point and a useful building block to think about interactive analogical retrieval. Of course, the application of this model assumes that the knowledge of these biological systems is *a priori* encoded in the long-term memory of the designer, which is why it cannot stand alone as an explanation for online bio-inspiration seeking.

Interactive Analogical Retrieval

The account of interactive analogical retrieval that is developed here rests on the following premises:

- Interactive analogical retrieval is sufficiently different from in-memory analogical retrieval because in-memory retrieval eschews interactions with the external environment.
- Interactive analogical retrieval is similar to garden-variety online information seeking tasks, in that they share the same substrate - external information environments.
- However, more crucially, interactive analogical retrieval is also sufficiently different from garden-variety information seeking tasks. Similar to its in-memory counterpart, interactive analogical retrieval plays a specific functional role in the overall analogy process. As such, interactive analogical retrieval is subject to the same *pressures of analogy* (Thagard *et. al.*, 1990) as in-memory analogical retrieval; these pressures – surface similarity (superficial features), deep similarity (abstract relations) and pragmatic relevance - originate in the higher-level analogy making process, of which retrieval is just one part. These pressures acting in

tandem is what distinguishes analogical retrieval and are not applicable to garden variety information seeking tasks.

- On the one hand, garden-variety online information seeking tasks are explained well within the framework of *Information foraging theory* (Pirolli, 2007), but it does not take into account the pressures of analogical retrieval. On the other hand, accounts of analogical retrieval (especially ARCS) take into account these pressures, but do not explain the retrieval process that involves interactions with the external information environment.

A high-level process account

Viewed as a black box, interactive analogical retrieval and in-memory retrieval are functionally similar in many respects. Both processes take as input a target problem and produce as output one or more source analogues. The difference between the two is the substrate on which the retrieval system operates: long-term memory, in the case of classical analogical retrieval, and external information environment, in the case of interactive analogical retrieval. But when we open the black box and look closely at the process of retrieval, the differences between the two are much more stark.

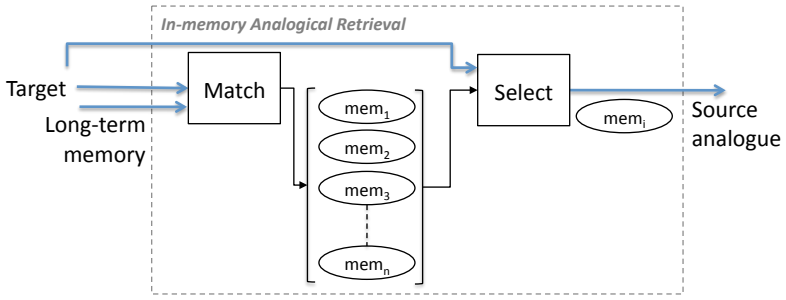


Figure 6.7: An abstract model of in-memory analogical retrieval

Generally speaking, in classical models of analogical retrieval (depicted in Figure 6.7) the target problem or aspects of it is used to probe the long-term memory in order to

access (and bring into short-term memory) or activate (to constitute the short-term memory) a set of memory elements (e.g., schemas, cases) that match the probe. Different models of retrieval make different commitments to the nature of probe, memory elements, matching criteria, etc. The matching stage is then followed by a selection stage, where a different set of criteria is applied on the matched memory elements to select one (or more) of those, which is regarded as the source analogue. Again, different models make different commitments to the selection criteria and how those get applied to the matched elements.

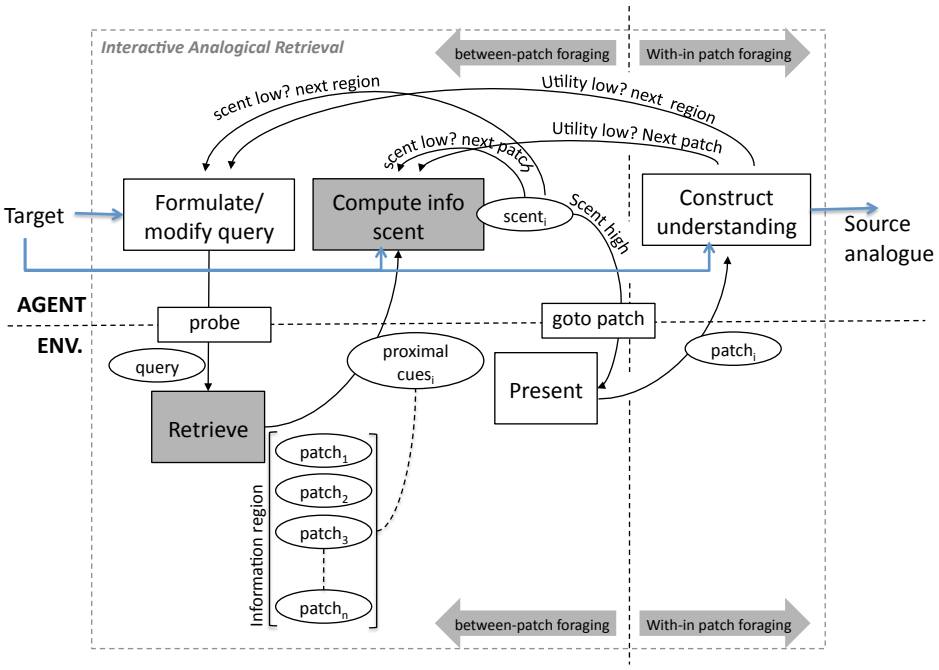


Figure 6.8: Interactive analogical retrieval model

Interactive analogical retrieval, on the other hand, is significantly different because the external information environment is a first-class object in the model, as depicted in Figure 6.8. The agent/environment divide is fundamental to interactive analogical retrieval model. But the scope of external environment in this work is restricted to online information environments only. Zooming into the interactive

analogical retrieval black box, we find that the high-level function is achieved by the same two processes that constitute the general information seeking behavior: *between-patch* and *within-patch* foraging processes.

Between-patch foraging using information scent in the context of interactive analogical retrieval as follows. Given a target problem or situation:

1. The analogist probes the environment by formulating and issuing a *query*. This query is context-dependent and draws upon the target problem.
2. In response, the environment retrieves and conveys an *information region* consisting of a set of *information patches* $\{(P_1,\{c_{11},c_{12},...\}),(P_2,\{c_{21},c_{22},...\}),...\}$, where P_i is an information patch and c_{ij}'s are the proximal cues associated with the patch P_i.
3. Forager perceives *information scent* of the patches, an estimation of how relevant different patches are to the target, based only on the visible proximal cues: $\{(P_1,s_1),(P_2,s_2),...\}$, where P_i is an information patch and s_i is the information scent that a forager associates with the patch P_i based on the match between the proximal cues and the target.
4. If the information scent of an information patch exceeds a certain threshold, it is considered relevant (high perceived utility). Therefore, the forager goes to that patch (by acting on the environment like clicking the associated hyperlink), at which point the environment presents the information patch to the forager. This initiates the *within-patch foraging* process.
5. If the scent does not exceed the threshold, it is considered irrelevant (low perceived utility), one of two things can happen as depicted in Figure 6.8: (1) the agent can stay within the same information region but loop back to Step 4 for processing the next patch in the region, or (2) the agent can abandon the current information region and loop back Step 1 in order to look for more fruitful regions.

6. Finally, there is *uncertainty* relationship between perceived information scent and the actual relevance of distal information patch – in some cases the scent might be high but the patch might turn out to be irrelevant and vice versa.

Once the forager picks up scent of a potentially useful information patch, the forager goes to that patch and starts consuming information in that patch, in what is termed as the within-patch foraging process. In the context of bio-inspiration seeking, this process involves comprehending the contents of an article and constructing a mental model of biological system(s) discussed in that article. In the within-patch foraging process, the agent is also simultaneously evaluating the actual utility of the patch by comparing/aligning/mapping the emerging mental model of the biological system against the target problem. In case of successful evaluation, the agent has obtained a source analogue. At any point, if this evaluation indicates a low utility of the current patch, the between-patch process is initiated. One of two things can happen when this transition occurs as depicted in Figure 6.8: (1) the agent can stay within the same information region but loop back to Step 4 (of between-patch foraging process above) for processing the next patch in the region, or (2) the agent can abandon the current information region and loop back Step 1 (of between-patch foraging process above) in order to look for more fruitful regions (see Figure 6.8).

There are two sub-processes in this analogical retrieval process where the pressures of analogical retrieval (semantic, structural, and pragmatic) might apply: "Retrieve" and "Compute information scent." These are depicted as boxes shaded in gray color in the Figure 6.8. That is because, these two sub-processes rely on the notion of similarity. The "Retrieve" process uses some notion of similarity that is built into the search algorithm in order to access information patches. The "Compute information scent" process computes the perceived utility of an information patch by computing the similarity between the target and the proximal cues associated with the patch. While the

information scent model provided in the general information foraging framework adequately explains the scent perception for non-analogy information seeking tasks, it has to be adapted in the present context such that it takes into account the three pressures of analogical retrieval.

Pressurized Information Scent Model (PRISM)

PRISM is a model of *information scent* perception in the context of interactive analogical retrieval. It explains how information scent is computed (perceived) based on the proximal cues found in the information environment using simultaneous satisfaction of a set of semantic, structural, and pragmatic constraints. This model is adapted from Analogical Retrieval by Constraint Satisfaction (ARCS) model developed by Thagard *et. al.* (1990), an in-memory model of analogical retrieval. ARCS is chosen because it is backed by psychological evidence and also because it lends itself well to this model. ARCS makes certain generally accepted assumptions about human memory that are retained in PRISM. In particular, this model assumes the presence of an organized store of associated concepts (associative semantic memory) to which representations of particular episodes are linked (episodic memory).

An analogist initiates the interactive analogical retrieval process with a particular target problem or situation in mind. PRISM assumes that the cognitive system of the analogist has represented the target problem in cognitive structures called *target schema*. The term schema has been used in the psychological and AI literature to mean many different things. For our purposes, a schema is defined as an explicit, declaratively-represented mental construct representing either an encountered or expected aspect of the world (Turner, 1994). A schema can be either learnt from experience or obtained from socio-cultural environment. For instance, Figures 6.9a and 6.9b depict the content of novice and expert target schemas of the problem of seawater desalination respectively.

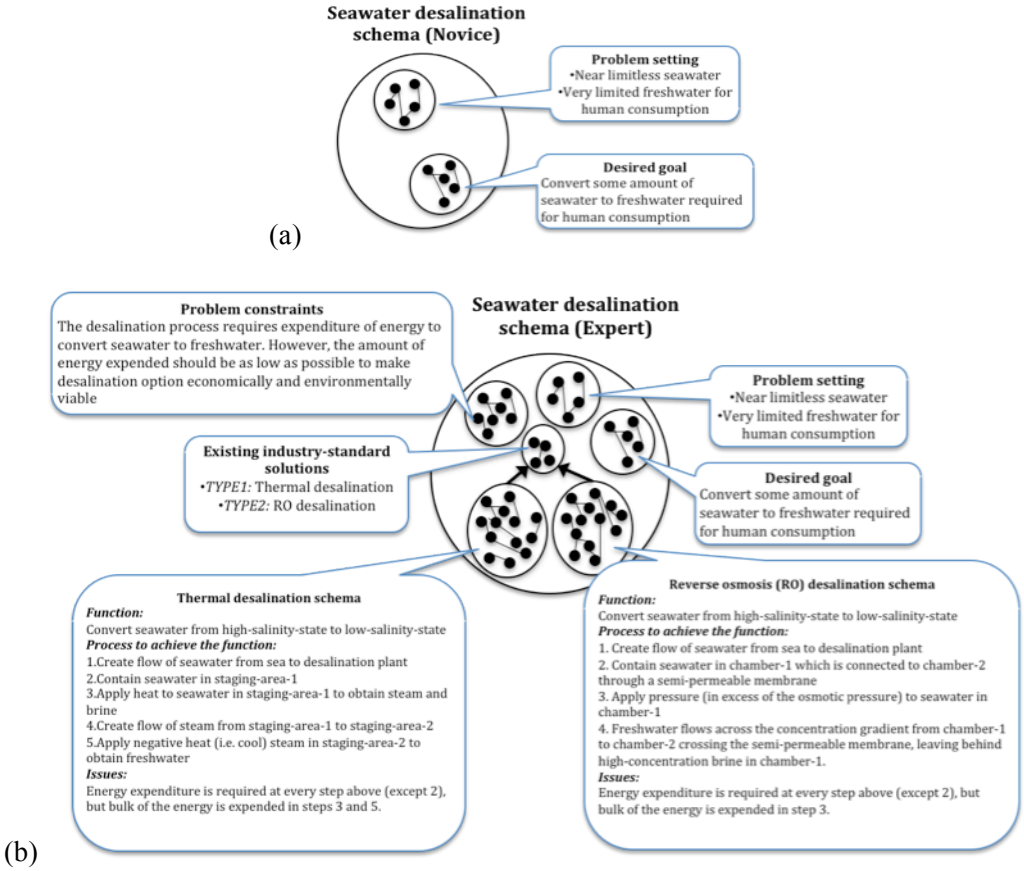


Figure 6.9: Hypothetical Target schemas of (a) novice, (b) expert

With a target problem in mind, the analogist forages the information environment for source analogues. During the between-patch foraging process, the analogist encounters a set of information patches with associated proximal cues as shown in Figure 6.10a. PRISM assumes that the goal of the analogist is to perceive (calculate) the information scent of each patch based on the proximal cues associated with that patch. The information scent of a patch will then allow the analogist to make judgment about the utility of going to that patch. This in turn allows the analogist to navigate the set of encountered information patches in the order of highest to lowest expected utility.

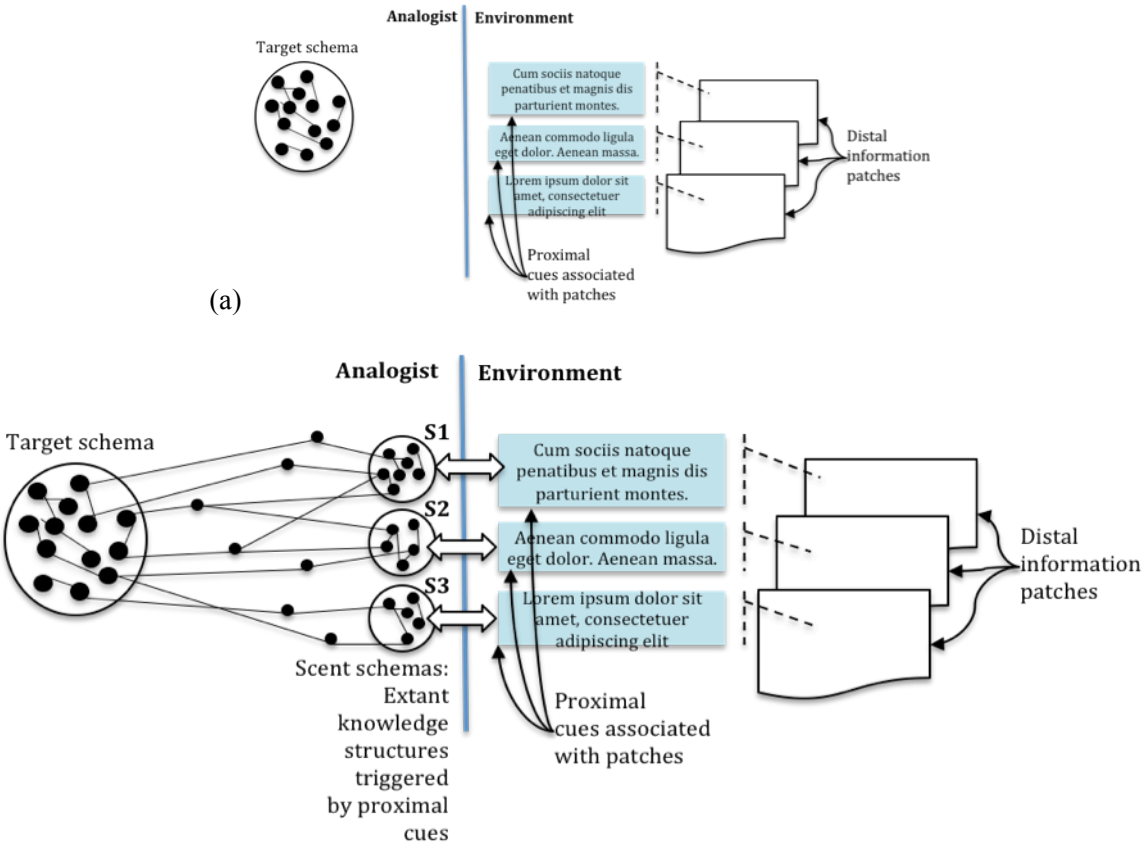


Figure 6.10: Scent perception in PRISM

When the analogist encounters proximal cues in the environment, PRISM assumes that the cognitive system of the analogist will represent those cues in cognitive structures called *scent schemas* as depicted in Figure 6.10b.

Given the target schema and scent schemas, PRISM computes the analogical similarity between the target and scent schemas in four stages, in a manner very similar to the original ARCS model.

In order to understand what those stages are, we have to make some minimal commitments about the knowledge representation of schemas and what those dots in Figure 6.10 mean. Let us assume that the conceptual structures representing these dots consist of propositions in predicate calculus. For instance, Figure 6.11 provides a very simple illustration of a target schema (T1) consisting of two propositions (P1-1 and P1-

2), and two scent schemas (S1 and S2) consisting of two propositions each (S1-1, S1-2, and S2-1 and S2-2, respectively). Let us also assume that the concepts *A* and *M* are semantically similar, and likewise concepts *B* and *N* are semantically similar; for instance, *A(a, b)* could represent *Regulate(kidney, potassium_ions)*, *M(m, n)* could represent *ControlProduction(pituitary, estrogen)*, *B(b, a)* could represent *IsSecretedBy(erythropoietin, kidney)*, and *N(n, m)* could represent *IsReleasedBy(hypothalamic_hormones, pituitary)*. Let us further assume that not all dots are equally important in the current context and that *A(a, b)* is more important than others.

The first stage: Initial network setup

Using information about the semantic similarity of predicates, the model creates a constraint network representing possible correspondences between concepts, predicates, relationships, and schemas as shown in Figure 6.11. This network is a connectionist network. Units representing correspondences are created and links between units are set up to indicate correspondences between the target and scents that support each other. The most important units are the ones that hypothesize that a scent schema is analogous to the target schema. Such units have names of the form TARGET=SCENT. (Here, "=" means "corresponds to," not identity). If the target is P1 and the scent is S1, then the unit created to represent a correspondence between them will be P1=S1. If P1-1 is a proposition in P1 that corresponds to proposition S1-1 in scent S1, then the unit P1-1=S1-1 that hypothesizes a correspondence between the propositions will have an excitatory link with the unit P1=S1. Moreover, units are created putting in correspondence the predicate and arguments of P1-1 with the predicate and arguments of S1-1, and these units receive excitatory links with the unit P1-1=S1-1. Excitatory links are also set up from a special semantic unit to predicate-predicate units based on the degree of semantic similarity of the predicates. The special semantic unit, like the special pragmatic unit of State 3, is a

unit whose activation level is always kept at the maximum value of 1. Hence it serves to pump activation to all units that are linked to it.

Target schema	**Scent schemas**	
P1	S1	S2
P1-1: A(a, b)	S1-1: M(m,n)	S2-1: M(n,m)
P1-2: B(b,a)	S1-2: N(n,m)	S2-2: R(n,m)
Assume: A and M are semantically similar; B and N are semantically similar; A(a,b) is most important in this context.		

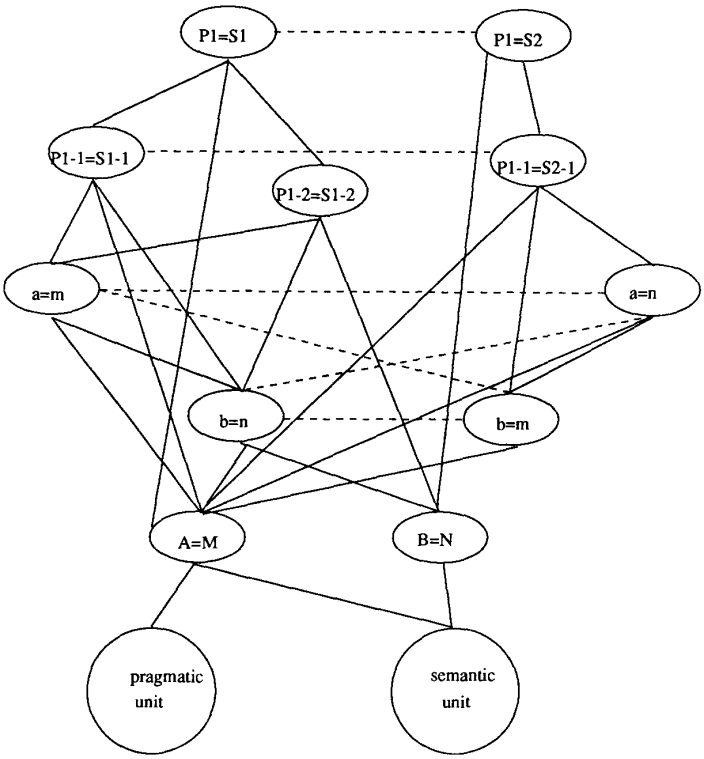


Figure 6.11: Setting up a multi-constraint network

The second stage: Adding inhibitory links

Inhibitory links are constructed between units representing incompatible hypotheses, for example, between P1=S1 and P1=S2. These make utility calculation competitive, in that choosing one scent will tend to suppress choosing of an alternative.

The third stage: Adding excitatory links

Pragmatic constraints are implemented by noting that certain elements (concepts, predicates, or propositions) are more *important* than others and that certain correspondences are *presumed* to hold in certain contexts. The information about presumed correspondences can originate from analogist's ossified knowledge. Excitatory links are set up from special pragmatic unit to all units involving "important" elements, and to all units representing "presumed" correspondences.

The fourth stage: Running the network

The constraint network constructed in stages 1 through 3 is run by setting the activation of all units to a minimal initial (random) level, except for the special semantic and pragmatic units for which activation is clamped at 1. Then the activation of each unit is updated by considering the activations of those units to which it has links. Cycles of activation adjustment continue until all units have reached asymptotic activation. The equation used for updating activation suggested in ARCS model: the activation of unit *j* on cycle *t + 1* is given by:

$$a_j(t+1) = a_j(t)(1-d) + enet_j(\max - a_j(t)) + inet_j(a_j(t) - \min). \qquad \text{(EQ.6.7)}$$

Here *d* is a decay parameter, $enet_j$ is the net excitatory input, and $inet_j$ is the net inhibitory input (a negative number), with minimum activation *min* = -1 and maximum activation *max* = 1. Inputs are determined by the equations:

$$enet_j = \sum_i w_{ij} o_i(t) \qquad \text{for } w_{ij} > 0; \qquad \text{(EQ.6.8)}$$

$$inet_j = \sum_i w_{ij} o_i(t) \qquad \text{for } w_{ij} < 0. \qquad \text{(EQ.6.9)}$$

Here, $o_i(t)$ is the output of unit I on cycle t, set by:

$$o_i(t) = \max(a_i(t), 0).$$

Updating the constraint network continues until all units have reaches asymptote, that is, a cycle is reached at which the activation change of each unit is less than a specified value, typically a low number (e.g., 0.001). For more fine-grained details about setting up the activation network, running such a network, computational complexity, etc. refer to Thagard et. al. (1990).

Analogical similarity

When the network settles, the analogical similarity between the target schema, T, and a particular scent schema, S_i, is equal to the activation value of the unit T=S_i in the constraint network. Higher the activation accumulated by the unit T=S_i the more similar is the scent schema, S_i, to the target, T.

Scent of an information patch

The information scent of a particular information patch, IP_i, which is associated with a set of proximal cues, $\{C_{ij}\}$, is equal to the analogical similarity between the scent schema, S_i, obtained from $\{C_{ij}\}$, and the target schema, T.

The dynamics of analogical similarity calculation and, as a consequence, the perception of information scent in PRISM is adaptive to the different knowledge conditions of target and scent schemas. One condition that can be identified is *sparse scent* schema condition: refers to scent schemas containing few isolated concepts and relationships. See Figure 6.12(a). Under this condition, since there is not a lot of higher-order structure in the knowledge, the computed information scent would mostly depend on the semantic similarity between the target and source schemas. Under this knowledge condition:

- If the target and scent schemas were from the same domain (implying significant semantic overlap), then the information scent would be strong.
- If the target and scent schemas were from different domains (implying low semantic overlap), then the information scent would be weak.

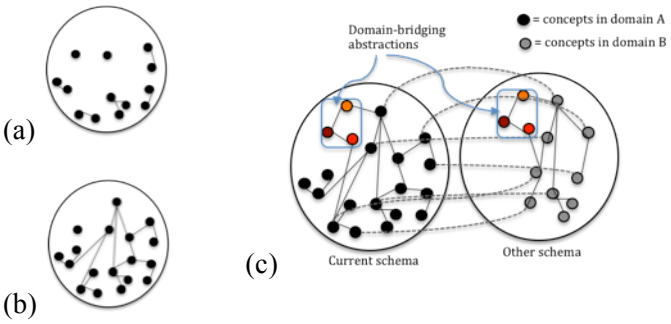


Figure 6.12: Knowledge conditions of scent schema

The second condition is *well-connected schema* condition: this refers to scent schemas containing not only concepts and relationships but also systems of abstract higher-order relationships that connect and organize lower-order relationships and concepts. See Figure 6.12(b). Under this condition, since there is structure in the knowledge, the computed information scent would depend on both semantic and structural similarity. Under this knowledge condition:

- If the target and scent schemas were from the same domain (implying significant semantic overlap) and the structural similarity was high, the information scent would be strong.
- If the target and scent schemas were from the same domain (implying significant semantic overlap) and the structural similarity was low, the information scent would be weak.

- If the target and scent schemas were from different domains (implying low semantic overlap), then the information scent would be weak[1].

Finally, the third condition that can be identified is *well-connected and shared abstraction schema* condition: well-connected schemas can be further classified based on the *encoding specificity*, which refers to how domain specific or general the content of the schemas are. In other words, the encoding specificity is related to the presence or absence of domain general abstractions as part of the higher-order content of the schemas. Research on memory, knowledge representation, and reasoning indicate that people learn domain general (shared) abstractions which allow them to see commonality between two or more schemas that differ only in specifics but share a deeper commonality. This essential idea is known by different names in different schools of analogy research, e.g., *thematic organization packets* (Schank, 1982), *thematic abstraction units* (Dyer, 1983), etc. I choose to refer to such abstractions as simply as shared abstractions. See Figure 6.12(c). Under this knowledge condition:

- If the target and scent schemas were from the same domain, the information scent would be same as in the simple well-connected knowledge condition discussed above.
- If the target and scent schemas were from different domains, then the domain-bridging abstractions will permit the construction of the constraint network (e.g., Figure 6.12(c)). Under such circumstances:
    - If the structural similarity between the target and scent schemas were high, the information scent would be strong.

---

1 To set up mappings between the target and the source some amount of semantic overlap is necessary. In the network of Figure 6.11, the statements "A and M are semantically similar" and "B and N are semantically similar" provided the semantic overlap, enabling the construction of the network. If there is no semantic overlap in the first place, then there is no network, and as a result structural similarity cannot be obtained.

- If the structural similarity between the target and scent schemas were low, the information scent would be weak.

Note that in the above discussion, there is no mention of pragmatic similarity although we discussed that that was one of the pressures of analogical retrieval. That reason for the omission is that the role of pragmatic unit in PRISM (and ARCS) is to only assign relative weighting to the different relations in the network. In other words, not all relations are created equal. Structure-mapping theory assumes that all higher-order relations of a particular order are equally important, and that, by default, higher-order relations are more important than lower-order relations. ARCS, however, assumes that the relative importance of relations depend on the context of analogy making. Introduction of pragmatics into the mix does not introduce any new concepts or relations to the schemas, but changes the weights accorded to the various relations between concepts. The target and scent schemas are pragmatically similar to the extent that they accord the same weightage to the shared relationships.

Efficiency of online bio-inspiration seeking process

The efficiency of bio-inspiration seeking process can be measured in terms of the *rate* of finding information patches that contain analogous biological systems.

Let:

n = number of "correct" information patches found that contain analogous biological systems,

T = total foraging time during which n patches were encountered,

$rate$ = rate of bio-inspiration seeking process,

P_i = period between finding $(i-1)^{th}$ patch and i^{th} patch,

T_{B_i} = between-patch foraging time in the period P_i,

T_{W_i} = within-patch foraging time in the period P_i,

By definition:

$$rate = \frac{n}{T}$$

But $T = \sum_{i=1}^{n} P_i$ and $P_i = T_{B_i} + T_{W_i}$

Therefore:

$$rate = \frac{n}{\sum_{i=1}^{n} T_{B_i} + T_{W_i}} \qquad \text{(EQ.6.10)}$$

We can further expand on T_{B_i} and T_{W_i} as follows:

$$T_{B_i} = t_{b_i} + \left(\sum_{j=1}^{F_i^{POS}} t_{b_{ij}} \right) \qquad \text{(EQ.6.11)}$$

$$T_{W_i} = t_{w_i} + \left(\sum_{j=1}^{F_i^{POS}} t_{w_{ij}} \right) \qquad \text{(EQ.6.12)}$$

Where:

t_{b_i} = between-patch time for the "correct" information patch found in the period P_i,

t_{w_i} = within-patch time for the "correct" information patch found in the period P_i,

F_i^{POS} = number of "false positive" information patches found in the period P_i,

r_i = number of information regions visited in t_{b_i} ,

F_i^{NEG}= number of "false negative" information patches overlooked in t_{b_i} ,

$t_{b_{ij}}$ = between-patch time for the j^{th} "false positive" information patch in the period P_i,

$t_{w_{ij}}$= within-patch time for the j^{th} "false positive" information patch in the period P_i,

r_{ij} = number of information regions visited in $t_{b_{ij}}$,

F_{ij}^{NEG} = number of "false negative" information patches overlooked in $t_{b_{ij}}$,

α = (assumed) constant time required to attend to proximal cues and compute scent in an information region,

β = (assumed) constant opportunity cost (in terms of time) for overlooking a "false negative" information patch.

But:

$$t_{b_i} = r_i \cdot \alpha + F_i^{NEG} \cdot \beta,$$

$$t_{b_{ij}} = r_{ij} \cdot \alpha + F_{ij}^{NEG} \cdot \beta$$

Therefore:

$$T_{B_i} = \left(r_i \cdot \alpha + F_i^{NEG} \cdot \beta\right) + \left(\sum_{j=1}^{F_i^{POS}} r_{ij} \cdot \alpha + F_{ij}^{NEG} \cdot \beta\right) \qquad \text{(EQ.6.13)}$$

Substituting (EQ.6.12) and (EQ.6.13) in (EQ.6.10) we get:

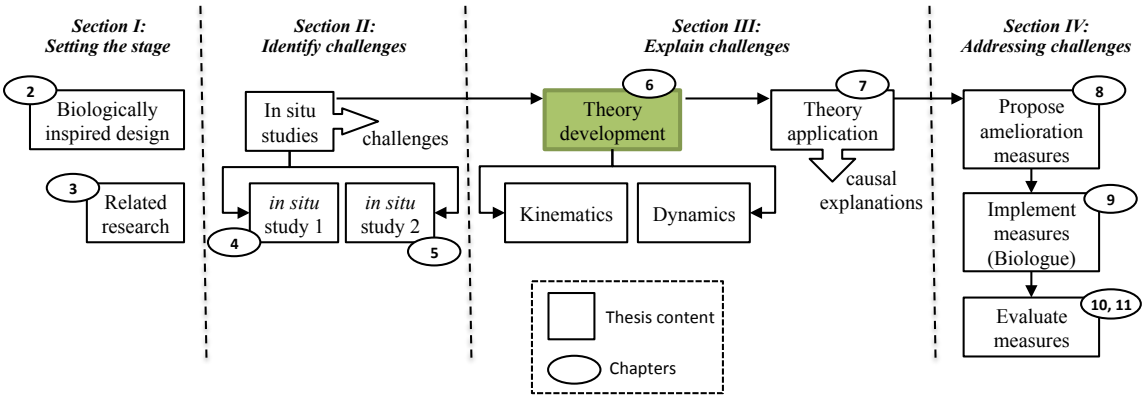
$$rate = \frac{n}{\sum_{i=1}^{n}\left(\left(r_i \cdot \alpha + F_i^{NEG} \cdot \beta + \sum_{j=1}^{F_i^{POS}} r_{ij} \cdot \alpha + F_{ij}^{NEG} \cdot \beta\right) + \left(t_{w_i} + \sum_{j=1}^{F_i^{POS}} t_{w_{ij}}\right)\right)}$$

Simplifying, we get:

$$rate = \frac{n}{\sum_{i=1}^{n}\left(r_i \cdot \alpha + F_i^{NEG} \cdot \beta + t_{w_i} + \sum_{j=1}^{F_i^{POS}} \left(r_{ij} \cdot \alpha + F_{ij}^{NEG} \cdot \beta + t_{w_{ij}}\right)\right)} \qquad \text{(EQ.6.14)}$$

Equation (EQ.6.14) shows that there are primarily 4 factors that influence the rate of bio-inspiration seeking: (1) number of information regions visited during between-patch foraging sub-process, (2) number of false positives, (3) number of false negatives, and (4) the within-patch foraging time (time spent consuming each article). Reducing any one of these four factors decreases the *period* between finding useful information patches (or, increases the *frequency* of finding useful information patches), thus increasing the *efficiency* of the bio-inspiration seeking process.

Chapter Summary



This chapter develops a theoretical understanding of online bio-inspiration seeking of designers engaged in biologically inspired design. This theory provides two kinds of accounts. First, the *kinematics* account provides a purely descriptive account of the phenomenon based on my analysis of the bio-inspiration seeking practice in the two *in situ* studies. Second, the *dynamics* account provides an explanatory account of the phenomenon in terms of its underlying causal processes or 'mechanisms.'

The dynamics account explains the online bio-inspiration seeking phenomenon in terms of an information-processing model that I developed called *Interactive Analogical Retrieval (IAR)*. The IAR model is a synthesis of two existing theoretical frameworks: one, a traditional cognitive model of analogical retrieval called Analogical Retrieval by Constraint Satisfaction (ARCS), and two, a human-information interaction theory which explains people's online information-seeking behavior called Information Foraging Theory.

The IAR model can be used to reason forwards from deliberate changes in the information environment to its observable effects on the online bio-inspiration seeking process of designers, or backwards from observed bio-inspiration seeking effects to the factors in the information environment causing those effects. In the next chapter we will

reason backwards with the IAR model and provide causal explanations for the three observed challenges associated with the online bio-inspiration seeking process.

CHAPTER 7

THEORY-BASED EXPLANATION OF THE CHALLENGES

Interactive Analogical Retrieval (IAR) model developed in the previous chapter provides the underlying process or 'mechanism' for online bio-inspiration seeking. In this chapter we will apply the IAR model in order to provide explanations for the identified challenges of online bio-inspiration seeking and identify the causes underlying those challenges.

Let us briefly revisit the three challenges here. First, designers experience a *low rate of encountering relevant information resources* in online environments that they normally rely on. Second, designers experience a *high rate of recognition errors*: they fail to recognize the true relevancy of the information resources that they encounter in those information environments. Third, designers experience *significant difficulty in comprehending information resources* that they recognized as being relevant and struggle to develop conceptual understanding of biological systems discussed therein.

An explanation for low rate of encountering relevant information resources

The *low rate of finding useful information* issue, where designers often go for long periods without finding a relevant information resource, can be localized to the loop highlighted in the IAR model shown in Figure 7.1. If this loop is executed too many times, then the number of information regions foraged will be high. From equation EQ.6.14. in Chapter 6, we can infer that as the number of information regions increase, the period increases or the frequency decreases. One reason for why foragers have to loop back is because the current information region does not contain patches that produce strong scents. This can be attributed to the retrieval or the access mechanism in the information environment.

In current common online information environments, keyword-based method of indexing and accessing of information resources is customarily employed, which support access to information resources based on literal similarity (word-for-word matching) while ignoring semantic-, structural- and pragmatic-similarity – the three pressures governing the process of analogical retrieval. This method does not support access to information resources based on the right kinds of things from a designer's perspective. As a result, each attempt at access can contain a large number of spurious information resources that are superficially related to the target problem. This resulting average low yield of information regions can result in an increase in the average number of information regions foraged and an increase in the average between-patch foraging time, resulting in the increase in the period between finding two useful information patches.

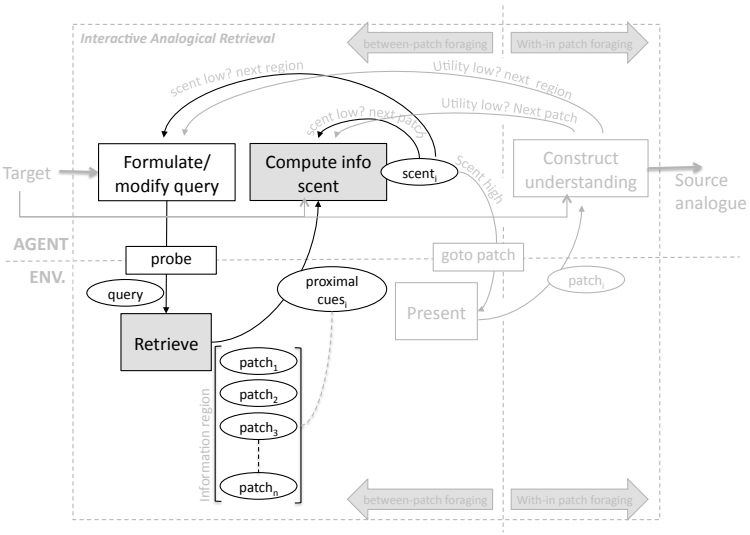


Figure 7.1: Localizing the issue of low find-frequency

This issue points towards the underlying problem well known in the CBR community, namely the *indexing problem* (Kolodner, 1993). Choosing the right indexing scheme is regarded as one of the most important factors in enabling the retrieval of relevant cases and preventing the retrieval of irrelevant cases. The three important qualities of a good indexing scheme are (Kolodner, 1993):

a) Indexing at storage time has to anticipate the vocabulary the analogist might use at retrieval time.
b) Indexing has to be by concepts and relations described a level of abstraction that is justified from a retriever's perspective.
c) Indexing has to anticipate the circumstances in which the retriever is likely to want to retrieve something (i.e., the task context in which the retrieval takes place).

All of the above are violated in when designers use current common information environments for retrieving their biological sources of inspiration. (a) is violated because the vocabulary of designers do not match with that of the biologists; by indexing the biology articles with keywords, the indexing favors the vocabulary of biologists, which is different from the vocabulary of the retriever (typically, an engineer). (b) is violated because the biology articles (especially, scholarly articles) discuss biological systems in terms of very specific issues and concepts, including extremely fine-grained analyses, and eschewing big-picture views and domain-neutral generalities; by indexing such articles with keywords, the indexing scheme does not favor a level of abstraction that is right from a designer's perspective which is not situated within the domain of biology. (c) is violated because when biology articles go online, they are indexed so that other biologists in the community can access and use them. They are *not* indexed anticipating that someday a designer might come along and use that article for creating a technology. Therefore, there is a mismatch in the anticipated circumstances and the actual circumstances with respect to indexing in the case of online bio-inspiration seeking.

Due to these violations, the accuracy of retrieval suffers, both in terms of *precision* and *recall* of retrieval. This has a direct bearing on the frequency with which one finds useful information in an information environment.

Therefore, the issue of *low rate of encountering useful information resources* issue can be traced to the current keyword-based methods of indexing and accessing

information resources in online information environments, which support access to information resources based on literal similarity (word-for-word matching) while ignoring semantic-, structural- and pragmatic-similarity – the three pressures governing the process of analogical retrieval.

An explanation for high rate of recognition errors

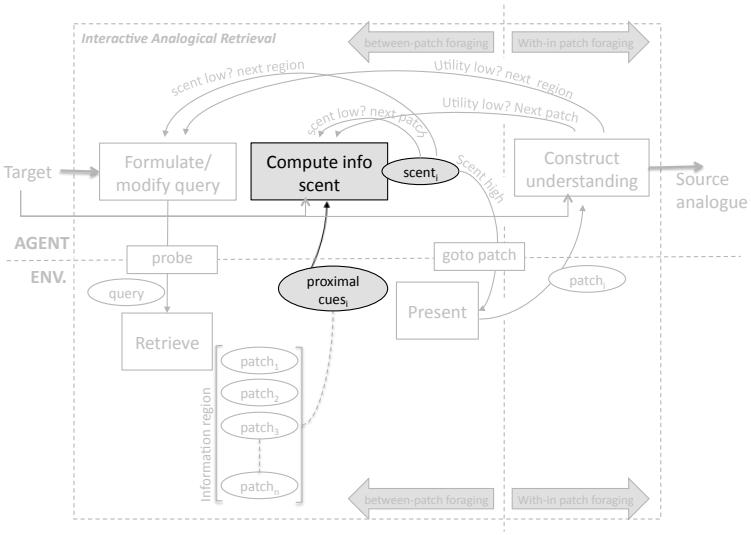


Figure 7.2: Localizing the issue of recognition errors

The *recognition error* issue can be localized to the information scent computation sub-process highlighted in IAR model shown in Figure 7.2. This issue is attributable to the nature of proximal cues that one encounters in customary online information environments – specifically, their lack of affordance for accurately perceiving the information scent of the resources they represent.

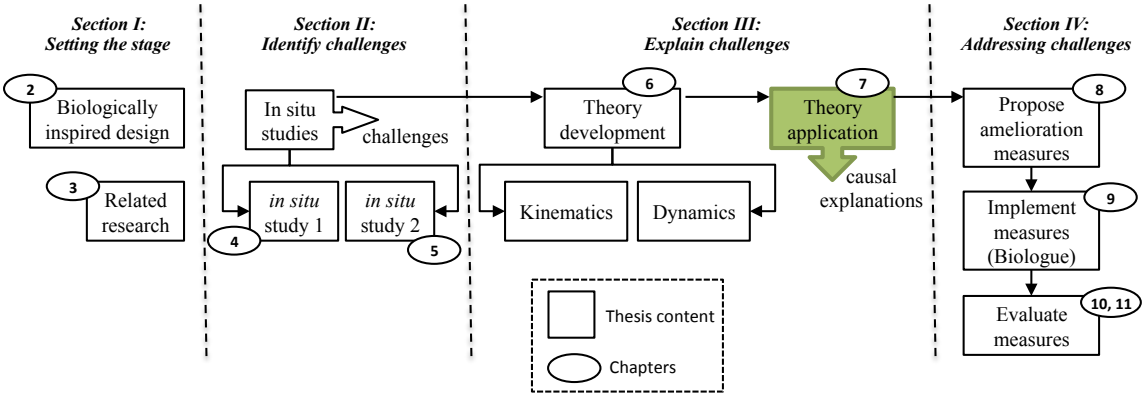
Accurately perceiving the scent of an information resource in the context of interactive analogical retrieval requires accurately judging the deeper analogical similarity between that target problem or situation and the source information that can be gleaned from the cues. But recall that the design of proximal cues customarily contains small snippets of information. With deep background information in biology, this may be enough for domain (biology) experts to infer the missing concepts and relationships

necessary to construct rich scent schemas. However, designers who are coming from engineering and who may not have the necessary background knowledge are more likely to be dealing with sparse scent schemas. According to the PRISM model presented in the last chapter, richer scent schemas afford computing the deep analogical similarity more accurately compared to sparse schemas (see previous chapter's discussion on scent computation under different knowledge conditions under the PRISM section). Therefore, the information scent computed by experts will be different from novices for the same given proximal cues. In light of this, novice designers are likely to make relevancy decisions based on superficial similarity as opposed to deep analogical similarity. This can lead to the rejection of information resources that contain structurally similar source analogues (false negatives) and/or selection of information resources that contain superficially or literally similar sources (false positives).

An explanation for difficulty in conceptual understanding

The issue of *conceptual understanding* can be attributed to the fact that existing biological information resources (especially scholarly articles) are usually created *by* and *for* biologists. They often do not contain the right kind of explanations for the uninitiated. The explanations, for instance, may not be at the right level of abstraction for non-biologists. The explanations may also leave a lot of information implicit, which constitute gaps in knowledge for non-biologists, requiring them to first develop the required expertise as part of the search process. The problem of retrieval is therefore often intertwined with the problem of learning in the context of seeking bio-inspiration. Scaffolding this process of learning appropriately can therefore significantly improve the efficiency of the interactive analogical retrieval process.

Chapter Summary



This chapter used the IAR model to localize the identified challenges of online bio-inspiration seeking and trace the underlying cause for those challenges.

SECTION IV: ADDRESSING THE CHALLENGES

Section Summary

The first research question addressed in this section is:

- ***RQ3: What measures can be taken to ameliorate the challenges of online bio-inspiration seeking?***

In order to answer the above research question, Chapter 8 proposes three hypotheses regarding the measures that can be taken in order to ameliorate the identified challenges. These three hypotheses are statements about the conditions in the external online information environment that make it favorable for engaging in the task of online bio-inspiration seeking.

Thesis 3: The identified challenges of online bio-inspiration seeking can be ameliorated by changing the conditions in the external online information environment that make it favorable for engaging in the task of online bio-inspiration seeking. In particular: (3a) The low rate of encountering relevant information resources can be ameliorated by semantically indexing and accessing biological information resources using concepts and relations derived from corresponding functional models (as opposed to the current dominant paradigm of keyword-based indexing and retrieval). (3b) The high rate of recognition errors can be ameliorated by enhancing proximal cues with visual overviews derived from corresponding functional models. (3c)The comprehension of biology documents challenge can be ameliorated by providing corresponding functional models as external representational aids to scaffold designers' understanding.

Based on the normative implications of the three hypotheses, I developed an online information-seeking environment called *Biologue* that is intended to better support the needs of designers engaged in online bio-inspiration seeking. Details about the design and implementation of Biologue can be found in Chapter 9 of this section.

The second research question addressed in this section is:

- ***RQ4: How does the presence of an online information environment with features grounded in the hypotheses associated with RQ3 change the dynamics of online bio-inspiration seeking in order to ameliorate the identified challenges?***

In order to address RQ4, I conducted a series of evaluation studies, including three experimental studies to evaluate RQ4.1, RQ4.2, and RQ4.3, respectively, and a pilot deployment study in which I deployed Biologue in Fall 2010 in the classroom context of ME/ISyE/MSE/PTFe/BIOL 4803, the same context that was initially used to conduct my *in situ* studies. Details of these studies can be found in Chapters 10, 11, and 12 under this section. The trends from these studies are encouraging and seem to suggest that measures put forth in RQ3 indeed change the dynamics of online bio-inspiration seeking in favor of ameliorating the identified challenges of online bio-inspiration seeking.

CHAPTER 8

PROPOSALS TO ADDRESS THE CHALLENGES

In the previous Chapter 7, I provided explanation for the challenges of online bio-inspiration seeking and identified the causes underlying those challenges. In this chapter I will propose measures that can be taken in order to ameliorate those challenges. These proposals are hypotheses about the conditions in the external online information environment that make it favorable for engaging in the task of online bio-inspiration seeking.

The First Proposal

Recall that the low find-frequency issue is attributable to the keyword-based method of indexing and accessing biological information resources. Alternate methods of indexing and accessing biology documents can help mitigate this difficulty. But the alternate methods of indexing must be consistent with the abstractions used by designers during their information seeking process. This leads to the following first hypothesis:

H1: In the context of online bio-inspiration seeking, indexing and accessing biological information resources using concepts and relations derived from their corresponding functional models will lead to higher rate of encountering relevant information resources when compared to keyword-based indexing and retrieval.

What is a functional model? The definition adopted by a functional model in this work is as follows. A functional model of a system (biological or human-engineered system) is a conceptual representation that specifies and organizes the functions of that system at multiple levels of abstraction and aggregation. At each level in this organization, a function specification is also accompanied by a specification of the causal processes that result in that system function.

What does "corresponding functional models" mean in H1? Let *D* be a biological information resource. Let *S* be a system discussed in that information resource. Let *F* be a functional model of S. Then, *F* is a corresponding functional model of *D* and vice versa. Because a biology article can discuss multiple biological systems, in theory there can be many corresponding functional models for the same document. Similarly, a biological system can be a subject of many biology articles. Therefore, one functional model can correspond to multiple biological information resources.

The Second Proposal

Recall that the recognition errors are attributable to the nature of the proximal cues and their affordance, or lack thereof, for accurately perceiving the information scent of distal information resources, suitably changing the design of proximal cues can mitigate this difficulty. This intuition leads to the following second hypothesis:

H2: In the context of online bio-inspiration seeking, enhancing proximal cues to include visual overviews derived from corresponding functional models will lead to lower rate of recognition errors when compared to proximal cues that do not include such overviews.

What does "corresponding functional model" mean in H2? Let *C* be a proximal cue in the environment. Let *D* be the distal biological information resource represented by *C*. Let *S* be a system discussed in that information resource *D*. Let *F* be a functional model of S. Then, *F* is a corresponding functional model of *C* and vice versa.

The Third Proposal

Recall that the understanding challenge is associated with the difficulty of learning the workings of biological systems from information resources containing explanations that are not suitable for the non-biologist designers. The third hypothesis posits that:

H3: In the context of online bio-inspiration seeking, supplying biological information resources along with their corresponding functional models as external representational aids will lead to deeper understanding when compared to supplying the information resources alone.

Rationale Behind the Proposals

The origin of the aforementioned hypotheses lies in thinking about the notion of similarity between technological systems (the targets of BID) and biological systems (the sources of BID). For an analogy from a source biological case to a target engineering problem to occur, there must be significant similarity between the target problem and the source case at one or more deeper levels of knowledge and/or inference. Further, for an analogy to work well, analogical transfer must take into account both the deep similarities and the dissimilarities between the target and the source; that is, knowledge of the deep similarity and dissimilarity entails inferences not only about what to transfer but also about what not to transfer. This leads to the following question: what are the contents of deep similarity between biological and engineering systems?

Of course, biological and engineering systems are deeply similar in underlying laws of physics and equations of mathematics. However, while any design must obey physical laws and mathematical equations, knowledge of physical laws and mathematical equations by and of itself does not necessarily lead to generation of novel designs. Building from some of the previous work on analogical design (Bhatta & Goel 1996, 1997, Goel 1992, 1997, Goel & Bhatta 2004, Goel, Bhatta & Stroulia 1997, Yaner & Goel 2006, 2007a, 2007b), it is posited here that the analogies in biologically inspired engineering design are based on *teleological similarity* - similarities in the functional abstractions and the causal processes or mechanisms that achieve those functions. Therefore, the rational for using functional models behind these proposals stem from the notion of analogical similarity operating at the level of teleology in this domain.

Previous work in systems thinking has led to the proposal of several different kinds of functional modeling schemes such as Functional Representation (Chandrasekaran 1994; Chandrasekaran & Josephson 2004; Sembugamoorthy and Chandrasekaran 1986), Function- Behavior-Structure (Gero 1990; Gero, Tham & Lee 1992), Function-Behavior-State (Umeda et. al. 1990; Umeda et. al. 1997; Umeda & Tomiyama 1996), and Structure-Behavior-Function (or SBF) (Bhatta & Goel 1994, 1997; Goel et. al. 1996; Goel & Bhatta 2004; Prabhakar & Goel 1997).

Although all these modeling schemes differ in various respects, they all share a common commitment to represent teleology using function as the principal organizing abstraction. Which one of these functional models works best for the aforementioned hypotheses is an open empirical question. But for the purposes of this dissertation, I will commit to one of them, namely, Structure-Behavior-Function (or SBF) modeling scheme.

This notion of teleological similarity between technological systems and biological systems underlies several aspects of the dynamics of the online bio-inspiration seeking process. Why?

Consider the first issue mentioned above – the low find-frequency issue. This issue arises because of the indexing problem discussed earlier. When designers search for relevant biology articles using search engines, the search engines are very helpful if they retrieve articles about systems that are teleologically similar to the target system being engineered. For instance, if the subject of design is desalination technology, search engines are most helpful if they retrieve articles about biological systems that address the abstract function of "removing or separating solutes from solutions" using some biological mechanism. Instead, when they are indexed and accessed based on keywords, this deeper similarity does not come through. As a result, current search engines favor more superficial or literal notions of similarity and they end up pulling out all articles in which desalination is literally mentioned. This typically results in the retrieval of articles that discuss previously engineered desalination technologies (literally similar to the target

design problem), or articles about biological systems where concepts such as salts and salt solutions are mentioned (superficially similar). If the searchers are lucky, some of these biological articles may also contain biological systems that are teleologically similar because they happen to contain the right keywords.

What this entails is that in order for search mechanisms to produce the right kind of search results, and thus improve the find frequency, the information resources have to be indexed and accessed using schemes that capture teleology of the systems discussed in those articles. Therefore, I hypothesize that changing the indexing scheme to one that is based on SBF representations can mitigate the issue of low find frequency.

Now consider the second issue – the issue of recognition errors during the search process. When designers use search engines to retrieve articles, they are presented with a list of surrogates of articles, some of which may be relevant but most of which may be irrelevant. Designers are then expected to look at these surrogates and make judgments about the potential utility of the distal articles those surrogates represent. I contend that teleological similarity plays a significant role in the accuracy of this decision making process – the closer the system in a distal article is to the target design, teleologically speaking, the higher the true utility of that article. The extent to which the surrogates afford the calculation of the teleological similarity is the extent to which designers' judgment of the distal articles would be accurate. Because we see that recognition errors are frequent, we can infer that the current state of affairs with the information surrogate design are less than ideal. But because it is posited that inclusion of SBF information may help with assessing the teleological similarity, I hypothesize that redesigning surrogates to include SBF information about the systems discussed in the distal articles can improve the true utility judgment, this improving the accuracy of information scent perception. One caveat, though: proximal cues are intended to aid *perception* of scent based on a quick scan of the information contained in them. Putting a lot of complex information and making the information seeking excessively deliberate, defeats the purpose of

proximal cues. Hence, I have claimed in my hypothesis that only an overview of SBF model should be included in the information surrogate redesign.

Finally, consider the third issue – the understanding construction issue. Once designers pick up the scent of an article, they go to that article and start consuming it. This refers to the process of reading and comprehending the contents of the article. The end goal of this act is to build a sufficiently rich mental model of the biological system discussed in the article to be able to (1) confirm the teleological similarity of that system to the target design, and (2) transfer the teleological knowledge of that system to solve the design problem. Again, the extent to which the content and form of biology articles afford the efficient building of the right kind of mental models to achieve (1) and (2) is the extent to which this issue is mitigated. But, we know that designers have trouble comprehending these articles in the context of BID. Therefore, we can infer that typical biology articles do not have the adequate affordance and that designers can use further scaffolding in this matter.

Previous work on engaging middle school students with SBF models in order to help them to come to a deeper understanding of complex systems has proven successful to some extent (Hmelo-Silver et al., 2008; Liu & Hmelo-Silver, 2009). We can learn from this work and try to bring similar learning benefits of SBF representation to designers in the present context. Hence I have proposed the third hypothesis that augmenting biology articles with SBF models of biological systems discussed in those articles can aid understanding-construction in the context of bio-inspiration seeking.

SBF Models of Biological Systems

Biological systems are complex systems. The literature provides many characterizations of complex systems (e.g., Dym 1994; French 1994; Forrester 1968; Hubka & Eder 1988; Pahl & Beitz 1996; Simon 1969, 1996; Suh 1990; Vogel 1998). For example, complex systems have emergent behaviors (e.g., French 1994; Vogel 1998),

complex systems are dynamical and contain feedback loops (Forrester 1968), complex systems have hierarchical structure, but are only nearly, not fully, functionally decomposable (Simon 1996), and so on. Narayanan (2007) characterizes a complex systems as follows: Complex systems exhibit hierarchical structures composed of subsystems and components; Subsystems and components exhibit natural behaviors or engineered functions; The subsystem/component behaviors causally influence other subsystems/components; The propagation of these causal influences creates chains of events in the operation of the overall system and gives rise to its overall behavior and function; These chains of events extend in temporal and spatial dimensions. We view this as a *minimal* characterization of complex systems. Note that these different characterizations of a complex system, including the minimal characterization, at least implicitly take a design stance towards and refer to its structure, behaviors, and functions.

Goel (1989), Goel & Stroulia (1996), and Prabhakar & Goel (1998) describe the view of complex systems that is espoused here. It builds on Simon's (1969, 1996) notion of a system's inner and outer environments. Let us first look at a characterization of a *bounded system* from the perspective of its inner environment as described by Goel & Stroulia (1996). The system has a structure, $\mathbb{S}$, consisting of its components and connections among the components, and a set of internal behaviors, $\mathbb{B}_{\text{internal}}$, consisting of its internal causal processes. The internal causal processes result in a set of output behaviors of the systems $\mathbb{B}_{\text{output}}$. An agent may ascribe a set of functions $\mathbb{F}$ to the system such that $\mathbb{F} \equiv \mathbb{B}_{\text{intended}} \subset \mathbb{B}_{\text{output}}$, that is, the functions are the intended behaviors of the system. As an illustrative example, consider the simple household flashlight. This flashlight has a structure $\mathbb{S}$ that consists of a battery, a bulb, a switch, all connected in series. The flashlight also has internal causal processes, $\mathbb{B}_{\text{internal}}$. These internal behaviors of the flashlight result in its output behaviors $\mathbb{B}_{\text{output}}$ such as production of light and production of heat. An agent may ascribe a single function $\mathbb{F}1$, generation of light, to the

flashlight, where $\mathbb{F}1$ is the intended behavior, $\mathbb{B}_{\text{intended}}$; that is a subset of the output behaviors $\mathbb{B}_{\text{output}}$.

Let us note several other features of the above characterization of bounded systems. A bounded system has only limited interaction with outer environment. In addition to its output behaviors, $\mathbb{B}_{\text{output}}$, a system's the interaction with its outer environment in this characterization is abstracted as a *stimulus* from the environment to the system (such as pressing of the switch on the flashlight) that perturbs the value of some variable in the system (e.g., changing the mode of the flashlight switch from open to closed) and activates the internal causal processes of the system. The internal causal processes $\mathbb{B}_{\text{internal}}$ are an intermediate level of abstraction between the structure $\mathbb{S}$ and functions $\mathbb{F}$ of the system. Function $\mathbb{F}$ is a teleological interpretation of the system. Thus, to continue our flashlight example, while in one context, an agent may view generation of light as its intended output behavior and thus its function, in another context, an agent may choose to ascribe generation of heat as a function to the flashlight. Note that not only are functions of a system a subset of its output behaviors, $\mathbb{F} \subset \mathbb{B}_{\text{output}}$, but also that any function $\mathbb{f} \in \mathbb{F}$ is an abstraction of a some internal causal process $\mathbb{b}_{\text{internal}} \in \mathbb{B}_{\text{internal}}$. When an agent ascribes a specific teleological interpretation $\mathbb{f} \in \mathbb{F}$ to a system, the agent views the specific internal causal process $\mathbb{b}_{\text{internal}} \in \mathbb{B}_{\text{internal}}$ responsible for the f as a *teleological mechanism.*

A detailed description of SBF representation is beyond the scope of this section. Instead, I will present an overview of the representation framework and through an example illustrate how it can used to reasonably model a complex system. For a more details, please refer to the published work (Goel, Rugaber, & Vattam, 2009).

An SBF model of a complex natural or technological systems system (1) explicitly represents its structure [S] (i.e., its configuration of components and

connections), its functions [F], and its behaviors [B] (i.e. its internal causal processes that compose the functions of the components into the functions of the system), (2) uses functions as indices to organize knowledge of behaviors, (3) annotates state transitions in a behavior with causal explanation for it, and (4) organizes the knowledge in an F → B → F → B … → F(S) hierarchy, which captures functionality and causality at multiple levels of aggregation and abstraction.

The origin of the Structure Behavior Function (SBF) representation (Bhatta & Goel 1994, 1997; Goel et. al. 1996; Goel & Bhatta 2004; Prabhakar & Goel 1998) lies in Chandrasekaran's FR scheme (e.g., Goel & Chandrasekaran 1989; Chandrasekaran, Goel & Iwasaki 1993). In particular, our SBF models both combine FR with Bylander's component-substance ontology and primitive functions (Bylander 1991), and extend FR to support the inferences needed for automated design (Goel 1992; Goel & Chandrasekaran 1989, 1992; Goel, Bhatta & Stroulia 1997). SBF models share the main features of the FR scheme: (i) functions of systems are represented explicitly, (ii) functions act as indices into internal causal behaviors responsible for them, (iii) behaviors are represented as an ordered sequence of states, (iv) state transitions in a behavior are annotated by the causal explanations for them, (v) the causal explanations can be of several types, e.g., component function, structural relation, domain principle, another behavior, and (v) the component function explanations for transitions act as indices into functions at the next (lower) level of aggregation. SBF models also extend the FR scheme: (a) SBF models use a component-substance ontology of devices, which enables a more precise specification of states in a behavior or in a function, (b) SBF models use an ontology of primitive functions based on the component-substance ontology, which enables a more precise specification of state transitions in a behavior, (c) in SBF models, separate behaviors are constructed for substances and components, which makes for a more precise specification of behaviors as a whole, (d) the functions of systems are viewed as a subset of its output behaviors, and SBF models allow specification of all

output behaviors of a device, (e) the internal causal behaviors in an SBF model may branch and merge, and (f) the internal causal behaviors admit inverse causality and bi-directional causality. While (d), (e), and (f) above enhance the expressive power of FR, (a), (b), and (d) afford more precise and accurate inferences needed for analogical design.

Examples SBF models

In this section I present two models of biological systems, Transpiration and Intestinal Peristalsis, using the SBF framework. My purpose for presenting these two models is (a) to concretely illustrate that the SBF representation schema is capable of modeling complex systems in biology, (b) to provide an example of visualization of SBF models, and (c) to highlight SBF's capability towards abstracting very complex systems into comprehensible terms.

Both examples share the same visual syntax to their figures. In Structure Models, we represent structural components as nodes and physical connections between components as annotated arrows. The directionality of the arrow tells how the physical connection is formed. For example, in Fig. 8.1 there is a connection between Soil and Water. One would read this as, "Soil contains Water." Note that for these particular models we do not differentiate between structural components and substances.

In Function Models, there will always be two nodes, which represent states of the system, and a single arrow connecting the node labeled "Initial State" (the first state of the system) to the node labeled "Objective State" (the final state of the system). The arrow represents a transition between those two states and, only in the case of Function Models, is always annotated with the name of the function. Below the annotation is a small, floating node that provides a list of external stimuli that affect the system during the function.

Behavior Models are similar to Function Models in that nodes represent states and edges represent transitions between states, with the arrow pointing towards the next state

in the Behavior. The Initial State of the system always has a dark bold outline relative to the other states. Note that in Behavior Models, a particular transition may be annotated with many causal explanations of varying types.

Transpiration

The first system we will describe is a model of Transpiration. Transpiration is the process of leafy (vascular) plants transferring water from the soil via roots, to the air via leaves. First, we will look at the Structure Model (see Fig. 8.1). This Structure Model describes only those components and substances that are relevant to the function of Transpiration at the level of abstraction we are modeling. Certainly, there are many more structural components and substances involved in the day-to-day activities of a plant, but as we will see, we need not complicate our model by including them.

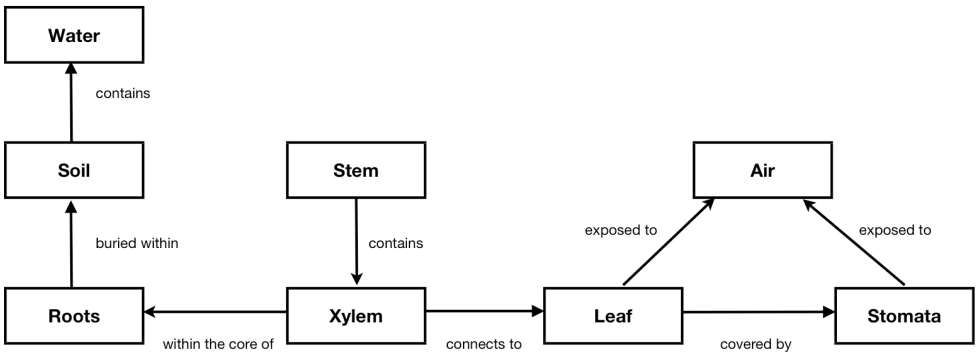


Figure 8.1: Structure model of transpiration system

Next, we will look at the Function Model (see Fig. 8.2). This model describes the initial state of the system, where liquid water is flowing in the soil, and the objective state of the system, where water vapor is escaping out of the plant into the air. Considering the natural language description of Transpiration provided above, these two states are reasonable, albeit highly abstracted, interpretations of the beginning and end states of the system.

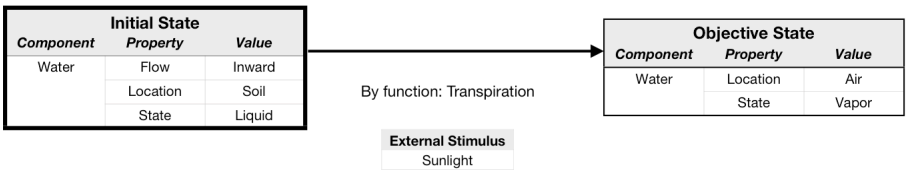


Figure 8.2: Function model of transpiration system

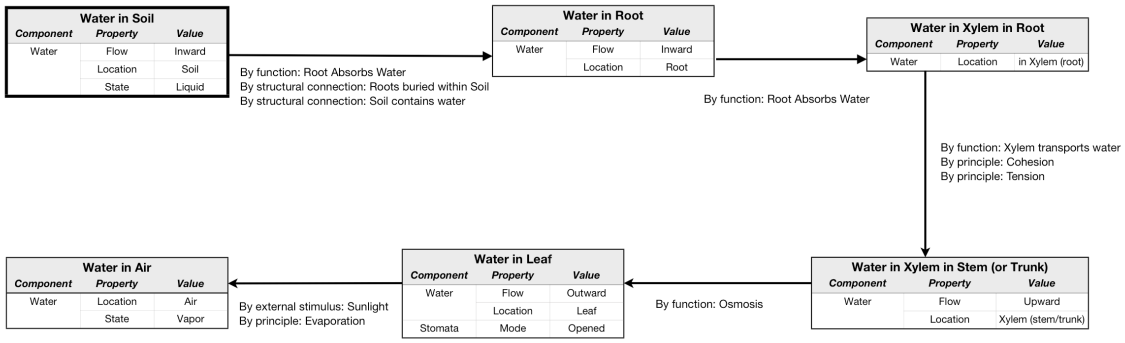


Figure 8.3: Behavior model of transpiration system

Finally, we look at the Behavior Model (see Fig. 8.3). The Behavior Model is an expanded version of the Function Model, so note that the Initial and Objective states are exactly the same as in Fig. 8.2. After the Initial State, where water is flowing in the soil, we transition to a state where water is flowing in the root. The reason this happens, as described by the annotations on the transition, is because the roots absorb water and are buried within the soil, which contains water. Next, the water moves from the roots into the xylem, which is contained within the roots. After this, the water is transported up the stem and then to the leaves. Once in the leaves, sunlight, something external to the system, heats up the liquid water in the leaf, converting it to water vapor and causing it to escape into the air.

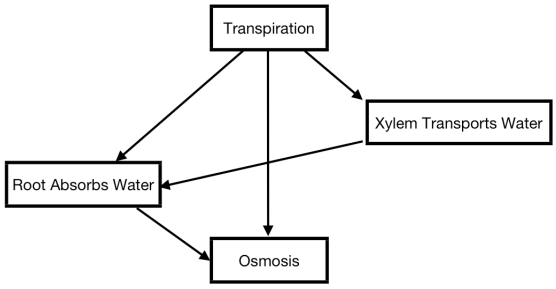


Figure 8.4: Function-sub-function hierarchy of the transpiration model

Note that the above model captures the transpiration system at just one level of the abstraction. But a more complete model will expand on the sub-functions that are part of the function-sub-function hierarchy in this model. This hierarchy is depicted in Fig. 8.4.

Intestinal Peristalsis

The next SBF model presented here is one describing Intestinal Peristalsis, which is the process by which chyme, a mushy substance that used to be food before being processed by the stomach, is transported through the human small intestines. The Structure Model (Fig. 8.5) shows that we will only be focusing on three components to describe this process: the chyme, the small intestines, and the smooth muscles that line the small intestines. Note that the connection between chyme and the smooth muscles is bi-directional, which means that both the chyme and the smooth muscles mutually press against each other.

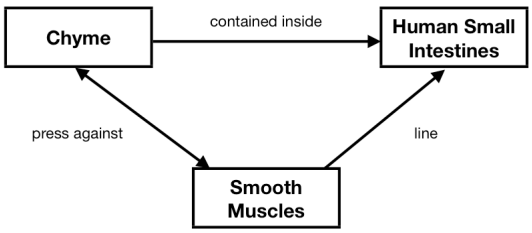


Figure 8.5: Structure model of intestinal peristalsis system

The Function Model (Fig. 8.6) of this system describes the initial state, where chyme is in the small intestines at some location X and the smooth muscles immediately behind the chyme are in a relaxed state. Then we transition to the final or objective state of the system, where chyme has progressed through the intestines by some distance Y. We do not use precise distance values here for two reasons. First, intestinal peristalsis works similarly through the small intestines, so we would like this model to be applicable for any location within the intestines. Second, we felt that at this level of abstraction it was only important to know that the chyme moved some distance, not the exact amount.

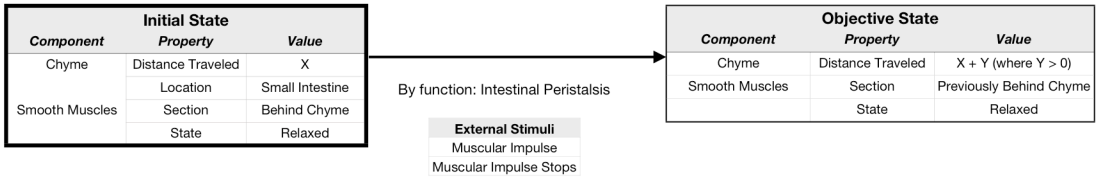


Figure 8.6: Function model of intestinal peristalsis system

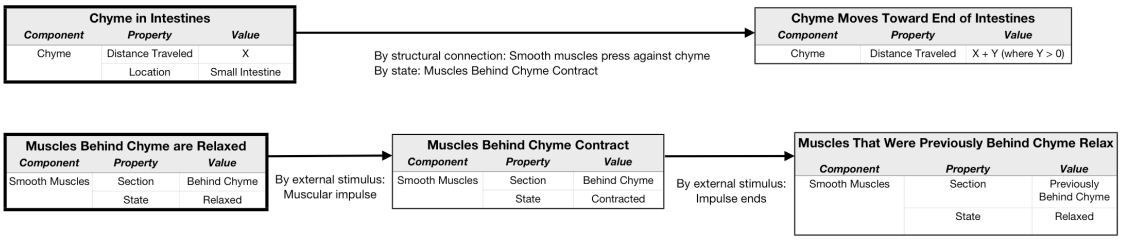


Figure 8.7: Behavior model of intestinal peristalsis system

The Behavior Model (Fig. 8.7) elaborates on the Function Model by providing intermediate states between the Initial and Objective state and by detailing the causal explanations between states. Observe that we have chosen to sub-divide the model into two separate causal chains, one for the chyme and one for the smooth muscles. Although this decision complicates the model, we did this because we wanted to make clear that the chyme does not move until the muscles behind the chyme contract. As in the Function Model, the behavior begins in a state where the chyme is at some location X and the

muscles behind the chyme are relaxed. Next, because of a muscular impulse (stemming from the nervous system) the muscles behind the chyme contract. Note that we did not progress further in the chyme's behavior because the second causal explanation (i.e., "By state: Muscles Behind Chyme Contract") has not yet been achieved. After this, the chyme is pushed forward by the contracted muscles some distance Y, and at some point in the future, the muscular impulse ends and the previously tense muscles relax.

To provide a sense of scale and complexity, we will now place the function of Intestinal Peristalsis in a greater context. Fig. 8.8 shows the Behavior Model for a system that describes how the human small intestines absorb nutrients, water, and lipids into blood and lymph fluid. Observe that Intestinal Peristalsis is simply one of many sub-functions in this Behavior Model (e.g., between "Pyloric Sphincter Opens" and "Chyme Moves to Duodenum"). It serves to provide a causal explanation for why chyme moves through the various sections of the human small intestines.

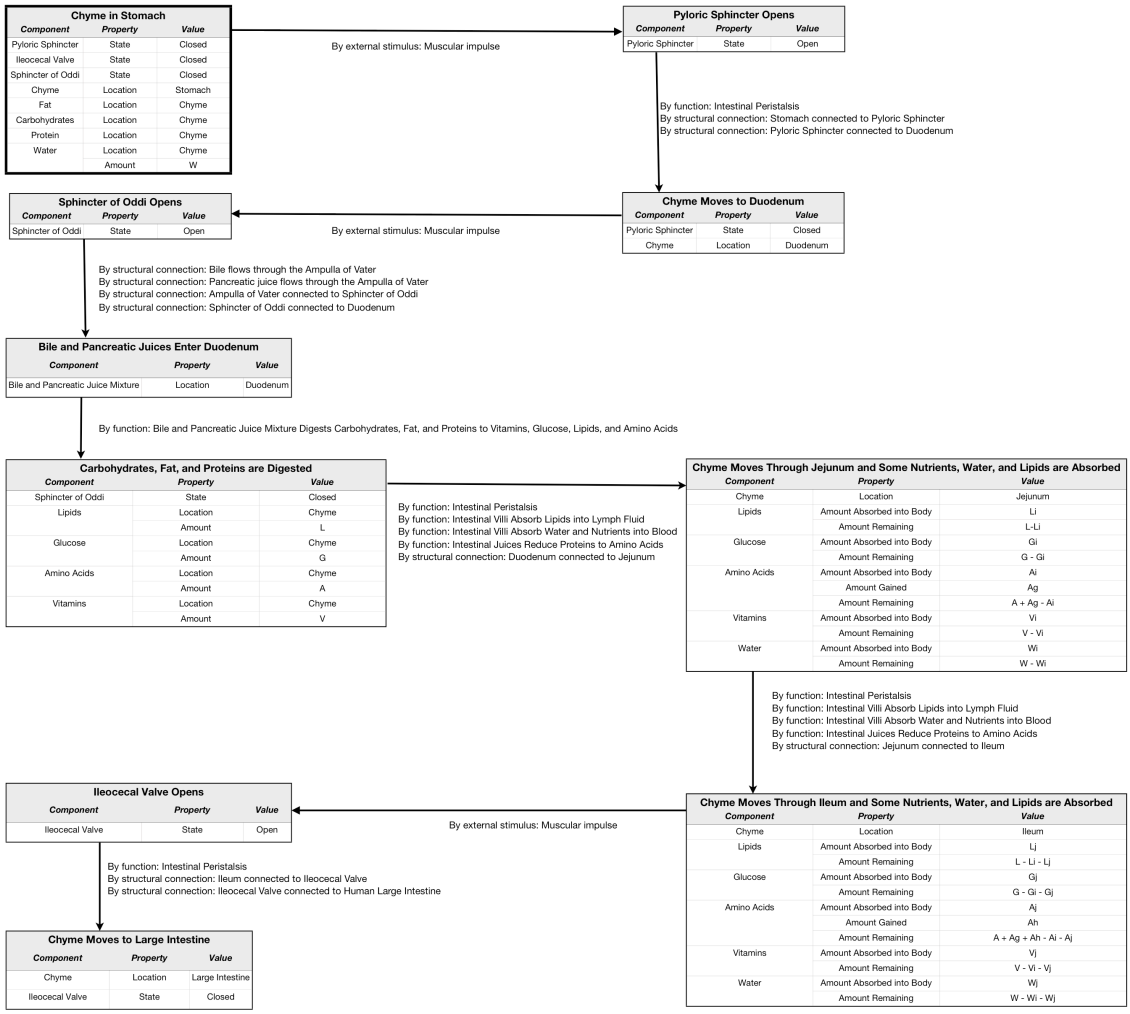


Figure 8.8: Behavior model of human small intestine absorb nutrients

It is also sometimes valuable to zoom out another level and just observe the function/sub-function relationships between SBF models. Fig. 8.9 illustrates the functional hierarchy for the small intestine model. Nodes represent functions and arrows between nodes show the function (base of the arrow) and sub-function (target of the arrow) relationship. Note that intestinal peristalsis is one of the sub-functions of this larger system.

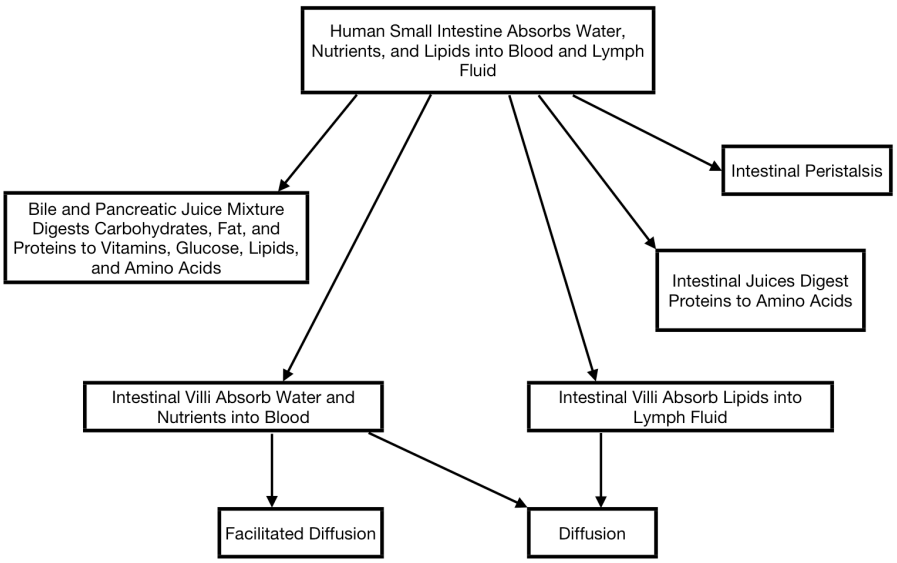
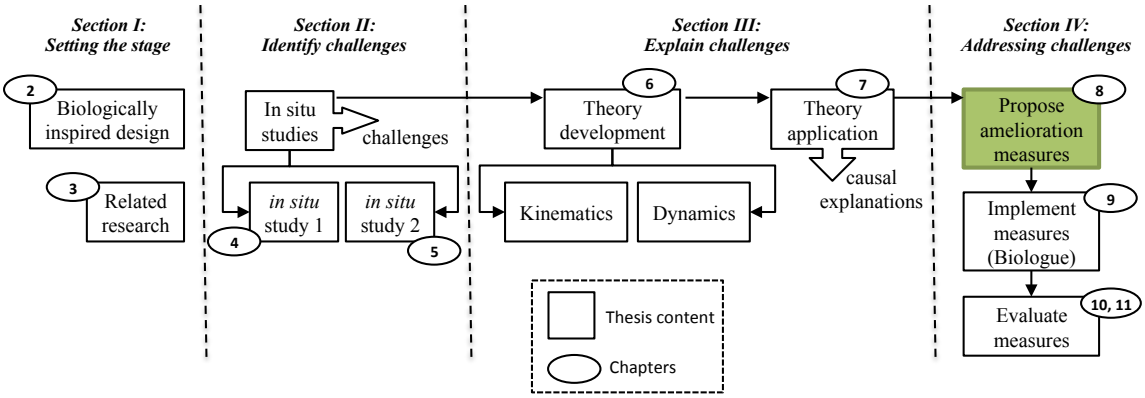


Fig. 8.9: Functional hierarchy for the small intestine model

Chapter Summary



This chapter proposes three hypotheses regarding the measures that can be taken in order to ameliorate the identified challenges. These three hypotheses are statements about the conditions in the external online information environment that make it favorable for engaging in the task of online bio-inspiration seeking.

The first hypothesis is a statement about indexing and accessing information resources in the external environment: low rate of encountering relevant information resources can be improved by semantically indexing and accessing biological information

resources using concepts and relations derived from corresponding functional models (as opposed to the current dominant paradigm of keyword-based indexing and retrieval).

The second statement is about the information content of proximal cues to improve the information scent accuracy: high rate of recognition errors can be lowered by enhancing proximal cues with visual overviews derived from corresponding functional models.

The third statement is about augmenting information resources with external representations to scaffold the process of comprehension: comprehension of biology documents can be improved by providing corresponding functional models as external representational aids to scaffold designers' understanding.

This chapter also discusses certain details associated with the hypotheses like functional models and provides examples of functional models of biological systems.

The proposals presented in this chapter will form the basis for the technology development effort of this dissertation which is discussed in the next chapter.

CHAPTER 9

BIOLOGUE

This chapter discusses the design and implementation of Biologue, a technology for aiding designers in their bio-inspiration seeking process. Design of Biologue embodies the three hypotheses mentioned in the previous chapter. Biologue's role is that of a *technology probe* (Hutchinson et al. 2003). Technology probes are research-specific systems that are intentionally simple and "underdesigned." Intended for use in field studies and/or controlled settings, these systems are intended to be technologically robust (e.g., they are not mockups or partially functional prototypes), and are instrumented to collect data. Technology probes aim to inspire and provoke discussion and ideas about future design possibilities rather than aim to be rollout ready for end use.

Design guidelines

The design of Biologue was motivated by the following guidelines.

1. Establish an online information seeking environment *by* and *for* biologically inspired design community such that:
    a. It has a growing repository of biology articles, and each article in its repository is augmented with SBF models of biological system(s) that are discussed in that article; when designers open an article to read, they also presented with its SBF model side-by-side, which acts as a scaffold for comprehending that article.
    b. All the articles in its repository are indexed and accessible using features derived from the SBF ontology such as functions, principles, operating environments, etc. such that the rate of finding useful articles is enhanced.

c. In the process of between-patch foraging in this environment, the proximal cues (information surrogates which are representative of distal biology articles) are presented to the seeker in such a way that they contain meaningful *overviews* that are based on their corresponding SBF models such that it is easy to distinguish useful articles from non-useful ones.

2. The establishment and maintenance of this environment should be *minimally invasive*: designers should not have to stray too far from their current practices in order to use Biologue. Biologue should be a natural extension to the socio-cultural environment depicted in the kinematics account of bio-inspiration seeking.
3. SBF models are structured representations. Manually obtaining structured representations requires knowledge-engineering effort. This knowledge-engineering effort should be distributed, community-driven, and emerge through social action.

The guidelines 1a, 1b, and 1c guides the design of Biologue such that it embodies the three hypotheses **H1**, **H2**, and **H3** discussed in the previous Chapter 8. While these three guidelines help establish the end goal, in terms of the properties that an ideal online information-seeking environment should possess for bio-inspiration seeking, guidelines 2 and 3 make commitments about the *means* or approach to getting there. No claims are made here about the goodness of this approach. The social approach to achieving such an environment may or may not be the most ideal one for getting there. It is merely one approach.

Socio-semantic foundations of Biologue

Social technology is characterized by an ecosystem of participation, where value is created by the aggregation of many individual user contributions. Semantic technology is an ecosystem of information content, where value is created by the integration of

structured information that is machine-readable. Socio-semantic technology, also known as *collective knowledge systems* (Gruber, 2007), synthesizes the strengths of these two kinds of technologies to create a new level of value that is both rich with human participation and powered by well-structured information.

Social technology is represented by a class of web applications in which user participation is the primary driver of value. The architecture of such systems is well described by Tim O'Riley (2012), who has fostered a community and media phenomenon around the banner of Web 2.0. Some of the important applications that represent this phenomenon include Wikipedia, Facebook, YouTube, Flickr, Del.icio.us, Technorati, etc. Discussion of the Social Web often use the phrase "collective intelligence" or "wisdom of crowds" to refer to the value created by the collective contributions of all the people writing articles on Wikipedia, sharing tagged photos on Flikr, sharing bookmarks on del.icio.us, or posting their personal blogs in the blogosphere. The excitement is understandable because the potential for knowledge sharing today is unmatched in history. Never before have so many people been connected by such an efficient, universal network. The result today is incredible breadth of information and diversity of perspective, and a culture of mass participation that sustains a massive source of publicly available content.

The role of semantic technology is to augment user-contributed data with well-structured information. While there are many ways to create value by aggregating user contributions today, there are few that go beyond summarizing or sorting the data. I see two major ways that semantic technology can significantly change the game. First, one can add value to user-contributed data by adding well-structured information. That is, semantic technologies can add structured data related to the content of user contributions in a form that enables more powerful computation. Second, the standards and infrastructure of semantic technologies can enable information sharing and computation *across* independent, heterogeneous social applications. By combining structured and

unstructured data, drawn from many applications across the Internet, semantic technology can provide a substrate for the discovery of new knowledge that is not contained in any one source, and the solution of problems that were not anticipated by the creators of individual applications.

The essential difference between non-semantic and semantic technologies is that structured data is exposed in structured ways. For example, the classic Web might have a document that mentions an organ, "Kidney." The conventional way to find this document on the Web is to search for the term "Kidney" in a search engine and manually pick out the pages that have something to do with the organ. The heuristics employed by today's keyword-based search engines for *inferring* what one means by the string "Kidney" is non-existent. Rather, they rely on the presence or absence of the string "Kidney," and rank those documents that contain the search string by *popularity.*

The semantic technology vision is to point to a representation of the entity, in this case an organ, rather than its surface manifestation. The heuristics for semantic search depends on conversation about how to represent things like organs (such as those specified in ontologies), and the availability of data which use these conventions. Such data is not available for most user contributions in social technology. To move to the next level of service, it would be nice to get the benefits of structured data from systems that give rise to the Social Web. There are three basic approaches to this: expose data that is already in the databases used to generate HTML pages, extract the data retrospectively from user contributions, and capture the data as people share their information.

The first approach is to expose the structured data that already underlies the unstructured web pages. An obvious technique is for the site builder, who is generating unstructured web pages from a database, to expose the structured information in those pages using standard formats. For instance, social web sites could expose their links to users as FOAF data, which is a Semantic Web ontology for representing personal contact information (Miller & Brickley, 2011). This, of course, requires the compliance of the

site builder, which means that it will not likely happen without a business motivation that benefits the site.

There are several promising techniques for the second approach, to extract structured data from unstructured user contributions (Auer & Lehmann, 2007; Mooney & Bunescu, 2005). It is possible to do a reasonable job at identifying people, companies, and other entities with proper names, products, instances of relations you are interested in (e.g., person joining a company) (Agichtein et. al., 2000; Cafarella et. al., 2007), or instances of questions being asked (Lita & Carbonell, 2004). There are also techniques for pulling out candidates to use as classes and relations, although these are a bit noisier than the directed pattern matching algorithms (e.g., Lin & Pantel, 2002; Pantel & Ravichandra, 2004). What is more interesting is that these techniques can be used to fold their results back into the data sources. That is, they can be used to augment the unstructured user data with structured data representing some of the entities and relationships mentioned in the text. For example, one could couple the structured data extracted from analyzing Wikipedia (such as that done by DBpedia), into tools that allow users to add structured data while they are entering wiki pages (such as Semantic MediaWiki). For instance, if a Wikipedia page mentions a book by its ISBN number, the link under the ISBN number could reference the book in structured databases of books and be used to call APIs (application programming interfaces) for obtaining it (Bizer et. al., 2007). More sophisticated examples for extracting references to named entities and factual assertions can also be applied. It is important to note that all these techniques require open data access and APIs to have a real impact on the Social Web.

The third approach is to capture structured data on the way into the system. The straightforward technique is to give users tools for structuring their data, such as ways of adding structured fields and making class hierarchies. This is naïve for the Social Web, since the users in this space are not there to create structured knowledge; they are there to have fun, connect with other people, promote their ideas, and share their experiences.

However, using interaction techniques such as *auto-completion*, it is possible to provide applications that give personal and social value to the individual in return for their using a tool that helps them add structure to their content (Gruber, 2007). Auto-completion is a feature provided by many web browsers, e-mail programs, search engine interfaces, source code editors, etc. It involves the program predicting a word or phrase that the user wants to type in without the user actually typing it in completely. This feature is effective when it is easy to predict the word being typed based on those already typed, such as when there are a limited number of possible or commonly used words (as is the case with e-mail programs, web browsers, or command line interpreters), or when editing text written in a highly-structured, easy-to-predict language (as in source code editors). It can also be very useful in text editors, when the prediction is based on a list of words in one or more languages. Auto-completion can result in a mix of structured and unstructured data, which has far more value when aggregated into collections. It is important to combine an auto-complete interface for soliciting structured data with motivations for providing this data. An interesting approach is to combine data entry with a social system that structures the behavior. For example, Luis von Ahn (2006) has created games in which people are rewarded for teaching the computer things such as what to label an image. The data structure is fairly simple: an entire image, or a well-defined region of the image, must be mapped to a word. The motivational structure of the game (try to match the label of other players) and the large number of players leads to quality of content.

An overview of Biologue design

Basically, Biologue is an online social bookmarking application for bio-inspired designers. It represents a fusion of social bookmarking services and traditional bibliographic management tools. It helps bio-inspired designers/researchers store, organize, share, and discover references to biology articles. Biologue belongs to the same

class of applications as other popular online reference management systems such as Connotea (www.connotea.org) and CiteULike (www.citeulike.org/).

Biologue is based on the principle of *social tagging* and is aimed to promote the sharing of biology articles among the designer community. In the same way that it is possible to catalog web pages (with Furl and del.icio.us) or photographs (with Flickr), designers can share information on biology articles with semantic tagging tools developed for the purpose of annotating and discovering articles. Tagging has rapidly become a common and popular practice on social websites. It allows people to easily annotate the content they publish or share with free-form keywords in order to make the content more easily browsable and discoverable by others, leading to a social component of tagging.

The functionality of Biologue can be separated into two basic tasks: *resource gathering* and *resource seeking*. In the resource-gathering task mode, one adds a reference to an article that he/she encounters during their research. In the current version, this reference information has to be manually added. But it is conceivable that in future versions, this can be done directly from the web browser, and for common online databases like PubMed, bibliographic data can be imported automatically.

Once a reference is added, one can manually add tags for annotating, organizing, and/or sharing an article reference. In most of the current online reference managements systems like CiteULike, this tagging keyword-based. While keyword-based tagging is a lightweight, agile and evolving way to annotate content, it can be efficiently combined with formal ontologies to make it more powerful, giving rise to *social-semantic tagging* (Passant, 2007; Fountopoulos, 2007; Hunter, 2009). There are only a few social bookmarking applications that support social-semantic tagging, including World Wide Web consortiums' Annotea project (http://www.w3.org/2001/Annotea/). Tagging in Biologue is a special case of social-semantic tagging. By virtue of the ontology used for tagging biology articles in Biologue, namely SBF ontology, the user-contributed

metadata that emerges result in the creation of (partial) conceptual models of biological systems discussed in those articles. In other words, as more and more people tag a particular reference, partially-structured SBF models of biological systems emerge in a piecewise, distributed fashion and get associated with that reference. Therefore, I refer to this special case of social-semantic tagging as *model-based tagging*. More details about model-based tagging will be provided in a separate section later on in this chapter.

New entries that are posted by users are added as *public* by default. Entries can also be added as private in which case they are only available to a specific user. Users can automatically share all their public entries with other users of Biologue. The semantic tags assigned to public entries contribute to the site-wide network of SBF models. All public references can also be searched and filtered by tags. In addition, Biologue provides *groups* that users can join. Groups are typically design teams working on a common bio-inspired design project.

In the resource-seeking task mode, Biologue provides an advanced search facility that includes searching based on features that are derived from the SBF ontology. Because these features match the kinds of features that designers naturally tend to use, we expect that this search mechanism can improve the rate of finding relevant articles in the context of biologically inspired design. Furthermore, the proximal cues (information surrogates) in Biologue contain overviews that are tailored to reduce recognition errors and help designers make more informed choices about which articles to pursue. Finally, when a particular article is selected for reading, Biologue provides a SBF-based visualization of its contents, which not only acts as a conceptual map but also as a scaffold for understanding its contents and building richer mental models of biological systems discussed within.

The architecture of Biologue

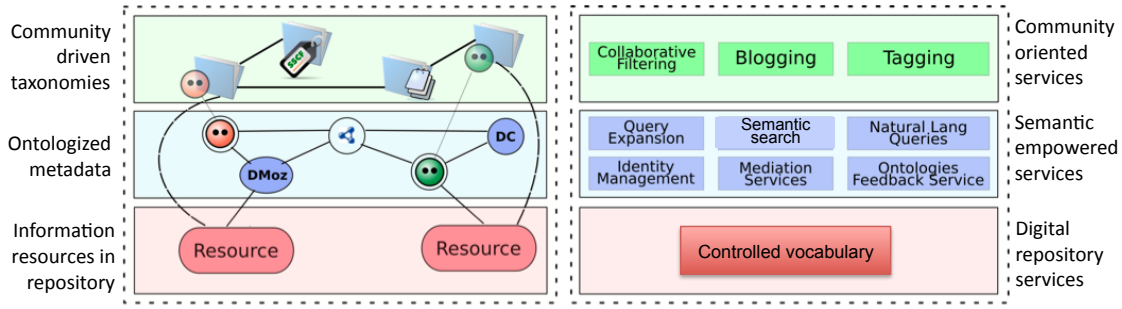


Figure 9.1: A generic socio-semantic architecture (adapted from (Kruk et. al., 2007), pp. 3)

A generic architecture of socio-semantic applications that is suitable for Biologue is presented in Figure 9.1 (Kruk et. al., 2007). It consists of a three-layered architecture, each layer enriching basic information gathered with semantic annotations and providing additional capabilities to browsing and searching.

The bottom layer handles typical tasks required from a digital objects repository, that is, keeps track of physical representation of resources, their structure and provenance. The bottom layer provides a service for a flexible and extendable electronic representation of objects; it is especially significant in expressing relations to other resources.

The middle layer lifts up legacy textual descriptions to a semantic level. It utilizes ontologies, an extensible description of the concepts and relationships, to represent concepts defined in legacy metadata formats (e.g., Dublin Core, MARC21 or BibTeX). The main advantages of the semantic layer are the services, which exploit machine-understandable, semantically rich relations between various kinds resources; they enhance the usability of information retrieval in the application or interoperability between different applications. An example service, a natural language query interface (Kruk et. al. 2006), may take advantage of a social network of users and creators, specified using FOAF metadata.

The top layer in the social semantic applications stack utilizes benefits from engaging community of users into annotating and filtering resources. In today's

applications the influence of user communities cannot be over-estimated; collaborative efforts in information sharing and management proved to be the right way to go and led to the success of many of the Social Web applications.

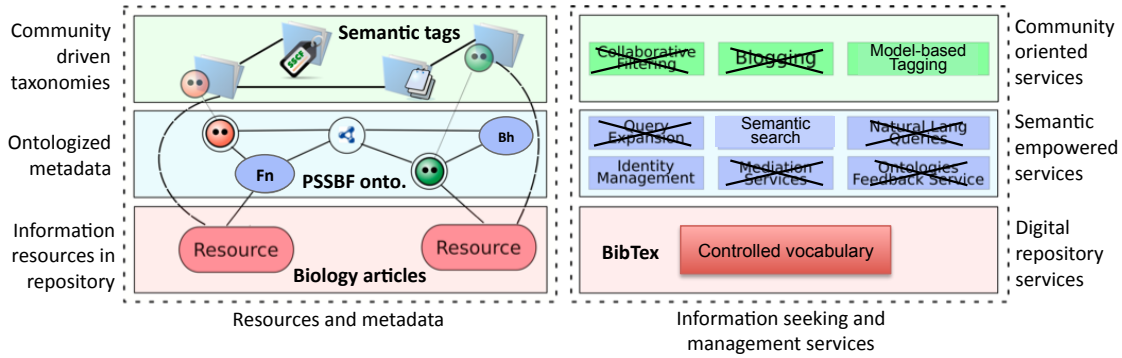


Figure 9.2: Current architecture of Biologue

Of course, Biologue in its current version does not incorporate all the elements of the generic architecture because it is not a full-fledged system. The elements of the socio-semantic architecture currently incorporated (and omitted) in Biologue are depicted in Figure 9.2.

On the *resources and metadata* side of the architecture, the bottom layer consists of a repository of biology articles that are contributed by the community. The middle layer consists of the ontologized metada that is derived from a variant of the SBF ontology called PSSBF (Partially Structured Structure Behavior Function) ontology. PSSBF ontology is discussed in a separate section a little later, but it is important to note that PSSBF is a less-strict version of the formal SBF representation, yet retains the essence of it. The upper layer consists of user-contributed tags that reference PSSBF metadata element, making them semantic tags.

On the *information seeking and management services* side of the architecture, the bottom layer consists of digital repository services for storing and organizing information about biology articles using the BibTex controlled vocabulary. The middle layer consists

of semantic empowered services. Currently, only the *semantic search* service and user *identity management* service is implemented in Bologue. The other services like Query expansion, Mediation services, etc. are not implemented (see the crossed-boxes in Figure 9.2). The top layer of Biologue's architecture consists of community-oriented services. Currently, tagging is the only service implemented in this layer. Other services like Collaborative Filtering, Blogging, etc. are reserved for future development should the need arise. But, the tagging service that is implemented in Biologue is special case of social tagging, which I refer to as *model-based tagging*, which is the subject of next section.

Model-based tagging

In order to make content more easily discoverable, social technologies use tagging, a technique where users can add free-form keywords, or *tags*, that act like subjects or categories for anything that they upload or wish to share. A tag is normally a single-word descriptor so punctuation marks are usually avoided, but some systems support phrases in quotation marks like 'global warming' and others use camelCase to distinguish between words.

One of the most popular tagging systems is the social bookmarking service del.icio.us, which allows one to store their favorite bookmarks on the Web via quick buttons in a browser (instead of locking them into a single desktop browser installation). Bookmarks saved in del.icio.us become accessible from anywhere and are normally public. After bookmarking a favorite URL, e.g. 'http://www.gatech.edu/', one can then add tags, e.g. 'university cool georgiaTech courses students'. Users can subscribe to other user's bookmarks, and bookmarks can be forwarded to other registered users using the custom 'for:username' tag syntax in del.icio.us. On the microblogging service Twitter, people have been using what are called 'hashtags' (i.e. tag keywords prefixed with the '#' or hash symbol) to annotate their microblog posts. While the use of hashtags began

several years ago, Twitter only added hyperlink support for these tags recently, such that clicking on a hash-tag brought one to a search service where related microblog posts using the same tag were shown.

While tags can be generally considered as a type of metadata, it is important to keep in mind that they are user-driven metadata. Tags focus on what a user considers as important regarding the way he or she wants to share information. The main advantage of tagging for end users is that one does not have to learn a pre-defined vocabulary scheme (such as a hierarchy or taxonomy) and one can use the keywords that fit exactly with his or her needs. Moreover, tags can be used for various purposes, and (Golder & Huberman 2006) have identified seven different functions that tags can play for end users, from topic definition to opinion forming and even self-reference. Marlow et al. (2006) also identified that in some cases, tags can be social elements that a user wants to emphasize, e.g. 'seen_in_concert.'

As tags are useful only when used in combination with the information resource they are related to, they are generally associated to tagging actions. A tagging action then represents the fact of assigning one or more keywords to online resources. Obviously, many tags can be assigned to the same resource, and on some services, different users can assign (the same or different) tags to the same resource. For example, in del.icio.us, a bookmark can be saved by several users, each of them being able to assign his or her own tags to the item. In order to simplify the tagging process, websites generally provide auto-completion features or automatically suggest tags, typically by analyzing tags already assigned by other users to the same resource.

From a theoretical point of view, a tagging action is often represented as a tri-partite relation between a User, a Resource, and a Tag as proposed amongst others by Mika (2005). Emerging from the use of tagging on a given platform, these actions lead to what is generally called a *folksonomy*, a term coined by Vander Wal (http://vanderwal.net/folksonomy.html) as a combination of the words 'folks' and

'taxonomy'. A folksonomy is hence a social, collaboratively-generated, open-ended, evolving and user-driven categorization scheme. Contrary to pre-defined classification schemes, users can use their own terms, which makes the folksonomy evolve quickly, based on the user's needs and benefiting from the 'culture of participation'. Websites that support tagging therefore benefit from the 'wisdom of the crowds' effect.

Information retrieval from tags and folksonomies is simply carried out using tag-based search engines. Folksonomies also provide a way to fluently navigate between various related tags and content, leading to serendipitous discovery of items. For example, users can generally navigate from one tagged item to the list of all items tagged with a similar tag, and so on. A popular visualization scheme for these tagging ecosystems is the use of tag clouds, where the highly-used tags are bigger (or bolder) than the other ones (similar to a weighted list in visual design). These tag clouds also give an overview of the main categories or topics discussed in the related community website.

There are several issues with free form tagging of content in social technologies. They include tag *ambiguity*, tag *heterogeneity*, and *lack of organization* among tags (Breslin et al., 2009). One way to address these issues with current social tagging system is through the introduction of semantic technology. In the past, folksonomies and ontologies have been regularly cited as opposite and exclusive means for managing and organizing information. A frequent point of view was to consider folksonomies as a bottom-up classification, while ontologies were seen as a centralized top-down approach. This way of thinking was also part of a larger set of opposing views between Web 2.0 and the Semantic Web. However, numerous recent works related to *social-semantic tagging* systems seem to suggest that this opposition may be unjustified and should not exist since these two fields are in fact complementary (and synergistic) paths towards enhancing the Web. Two separate approaches have been explored to synergistically combine social tagging and semantic technology: (1) one approach aims to define, mine or automatically link taxonomies or ontologies from existing folksonomies (e.g., Golder

and Huberman, 2006; Halpin et. al., 2006; Angeletou 2008), and (2) the other approach defines ontologies for tags and related objects to comply with (e.g., Gruber, 2007; Kim et al., 2007; Abel et. al., 2007). The borer between both is sometimes fuzzy since both approaches can be combined together.

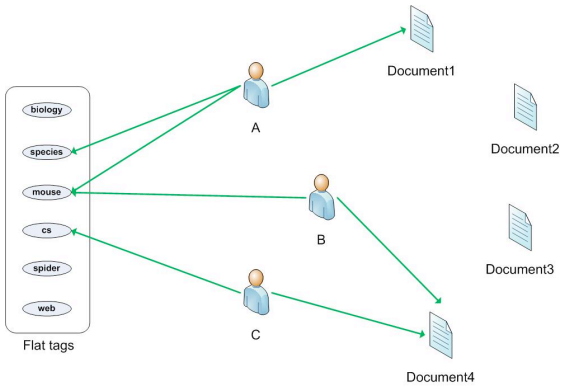


Figure 9.3: Flat tags in current social tagging systems (from Fountopoulos, 2007, pp. 12)

The tags in current social tagging systems are flat (pure), meaning they are not connected in any way by some types of relations between them. This scenario is depicted in Figure 9.3. Social-semantic tagging takes this further by introducing semantic relations between tags using some kind of ontology. For example, the RichTags system (Fountopoulos, 2007) uses the SKOS ontology (Miles & Brickley, 2005) as a model for expressing semantic relations between tags. This scenario is depicted in Figure 9.4a. A sample of the resulting semantic network of tags that emerges is depicted in Figure 9.4b.

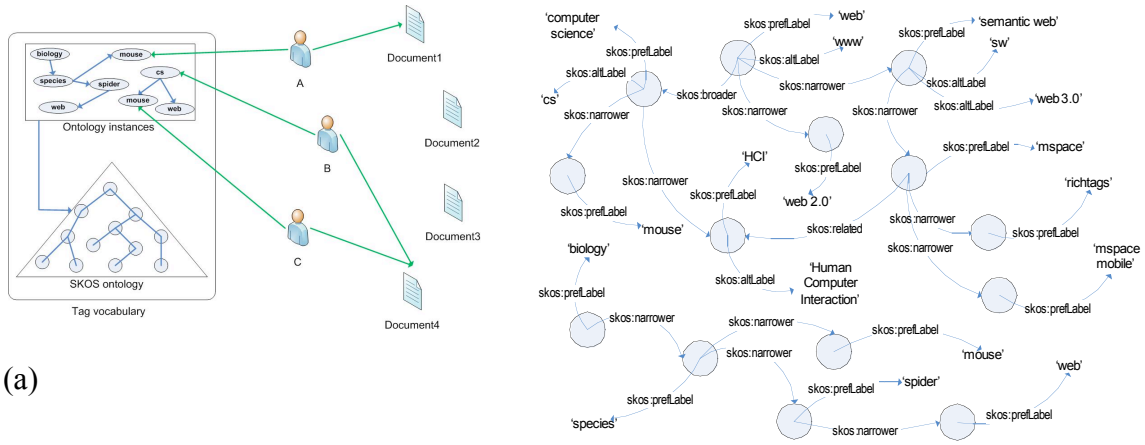


Figure 9.4: (a) Social semantic taggin scenaio in RichTags (from Fountopoulos, 2007, pp. 13), (b) a sample of the semantic network of tags that emerges in RichTags (from Fountopoulos, 2007, pp. 11)

Similar to RichTags, Biologue also takes the second approach of using predefined ontology to express semantic relations between tags. The ontology used by Biologue is the PSSBF ontology, which is discussed in the next section. By virtue of using this ontology, the social act of tagging articles in Biologue leads to the emergence of partially-structured SBF models of biological systems associated with those articles. Over time, parts of these models get reused and recombined. This is because of the auto-complete feature: as user starts typing a tag, an existing tag that has a similar beginning is suggested. If the user accepts the suggested tag, it gets reused along with all its existing relations. This eventually leads to the emergence of a site-wide network of partially-structured SBF models. This scenario is depicted in Figure 9.5.

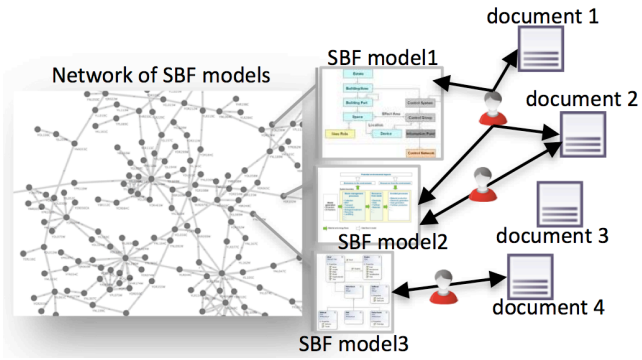


Figure 9.5: Model-based tagging scenario in Biologue

It is important to state here that model-based tagging represents just one of the means to establishing an online information environment that have the desired properties or qualities that favor online bio-inspiration seeking (as laid out in my hypotheses). At this stage, no claims can be made about the efficacy of model-based tagging technique for achieving this end goal and remains outside the scope of this thesis. The pilot deployment study discussed in Chapter 12 touches upon some of the efficacy issues.

PSSBF ontology

PSSBF is a modified version the *SBF ontology*, published in Goel, Rugaber & Vattam (2009). Modifications were made to the original SBF ontology for two reasons. First, the original SBF representation was catered to model human-engineered systems, but the representation here requires modeling biological systems. Therefore, new categories need to be introduced like operating environment and subject, verb and object properties of functions. Second, the original SBF representation was fully structured and representing behavior and structure model portions of the SBF model required heavy knowledge-engineering. In PSSBF ontology, a tradeoff has been made between knowledge-engineering effort and fully-structured representation, swapping purely symbolic representation of behaviors and structures for natural language descriptions of the same. As a result, PSSBF models of biological systems are partially-structured representations, containing a mix of symbolic and textual information. In the context of Biologue, the swapping of symbolic representation of behavior and structure for textual descriptions of the same are justified because we are not interested developing an AI system that is capable of engaging in causal reasoning about complex systems. SBF was originally developed for AI systems. Here, we are repurposing it for aiding humans. On the one hand, we cannot do away with symbolic representation of systems altogether because we need some structured information for conceptual indexing and such. On the

other hand, because we are not dealing with AI systems, we do not need a purely symbolic representation of systems as it is too expensive to knowledge-engineer. PSSBF tries to walk the middle path between these two limiting conditions without compromising the essence of SBF representation framework, which can be enumerated through the following 7 principles:

1. Explicit representation of function
2. Behavior is an intermediate level of abstraction between function and structure.
    a. A causal process is decomposed into states and state transitions.
    b. A behavior may be described with branches, merging, and iterations.
3. Functions are indices into behavior.
    a. A function is a state abstraction of a behavior in that it is associated with only the initial and objective states of the behavior.
4. State transitions in a behavior are annotated by causal explanations.
    a. There exists a set of causal explanation types.
5. Through "by function" causal explanations, one can compose a system-subsystem decomposition.
6. The system-subsystem decomposition bottoms-out in physical structures and domain principles.
7. A function interfaces with the external environment through external stimuli.

This section presents a top-down, syntax-oriented grammatical description of PSSBF. The notation used is a variant of BNF (International Organization for Standardization 1996) in which syntactic definitions are described using production rules in which the term being defined appears on the left of a separator (**:=**), and its definition appears on the right as a sequence of terminal and non-terminal symbols. *Terminal symbols* denote categories of atomic words in an SBF model. Important categories in SBF

are **STRINGS** and **INTEGERS**. Other textual names denote *non-terminal* units that are defined in other rules in the grammar.

Several operators are used in the syntactic definitions. Juxtaposition denotes catenation; **'|'** denotes alternative; **'[T]'** denotes optionality, where **T** is any string of symbols; **'{T}*'** and **'{T}+'** denote respectively any number of occurrences of **T** and any non-zero number of occurrences. Finally, **'// ...'** denotes a comment that proceeds from the slashes to the end of the line. Together, the following set of rules comprises an abstract syntax for PSSBF. The syntactic description is abstract because it avoids concrete details such as punctuation and keywords.

At the highest level, a PSSBF specification looks like the following:

PSSBFModel := INTEGER //ModelID
STRING //Model name
[STRING] //Description
StructureModel
FunctionModel
BehaviorModel

That is, a PSSBF specification comprises six parts, appearing consecutively: a unique modelID, a name, an optional description, specific submodels for structure, function, and behavior. These latter three constituents are now described in corresponding subsections.

The function model specification looks like the following:

FunctionModel := {Function}+
Function := INTEGER //Function Id
Name //Function name

[STRING] //Description

STRING //Subject

STRING //Verb

[STRING]* //Object(s) of the function

[STRING]* //Operating environment(s)

{Function}+ //Sub-functions of this function

INTEGER //BehaviorID of the behavior associated //with this function

INTEGER //StructureID of the structure associated //with this function

The function model consists of a set of functions. Each function is described by a name, description, a subject of the function, a verb of the function, and (optionally) the object(s) of the function. Additionally, a function may point to a set of operating environments this system operates in, a (optional) set of sub-functions of this particular function, a pointer to a behavior model which describes the behavior that accomplished this function, and a pointer to a structure model which describes the structures that participate in the function.

The structure model is a textual description of the structure of the system at a particular level in the function hierarchy. The specification is as follows:

StructureModel := INTEGER //StructureID

STRING //Structure model description

The behavior model is a textual description of the behavior of the system at a particular level in the function hierarchy. Additionally, a behavior model can point to a set of physical principles that underlie the behavior. The specification is as follows:

BehaviorModel := INTEGER //BehaviorID

STRING //Behavior model description

[Principle]* //Physical principles //underlying behavior

Principle := INTEGER //PrincipleID

STRING //Description of the principle

System description

In this section I will discuss the specifics of the implementation of Biologue. Every user of Biologue has to fist sign up for an account. Once an account is created and a personal profile has been established, they can login into Biologue as shown in Figure 9.6.

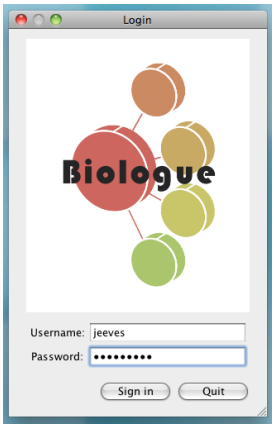


Figure 9.6: Logging into Biologue

Once a user logs in, they see their personal workspace, which contains a list references to the articles they have posted in Biologue. This is shown in Figure 9.7. If they have not posted any citations, their workspace will be empty.

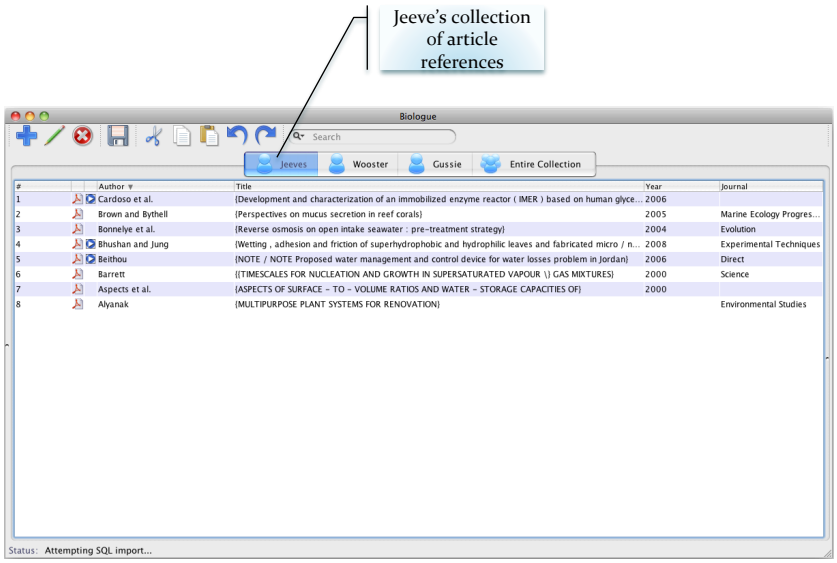


Figure 9.7: A snapshot of a user's personal workspace

Adding a new entry in the workspace is handled by the click of a button. Once a new reference is added, they can manually enter the bibliographic information about the article as shown in Figure 9.8. One can also avoid the manual entry by *importing* a reference in BibTeX format, at which point the bibliographic information is automatically added in Biologue. When a citation is posted, its status is *public* by default. That is, every user in Biologue can see this post f they search for it. However, its status can be manually set to private. They can also upload and attach a PDF version of the article if they have access to it. This raises several issues about copyright and the public disclosure of that uploaded article. These issues are beyond the scope of the current work because Biologue is merely a technology probe at this point and not rolled out for public use yet.

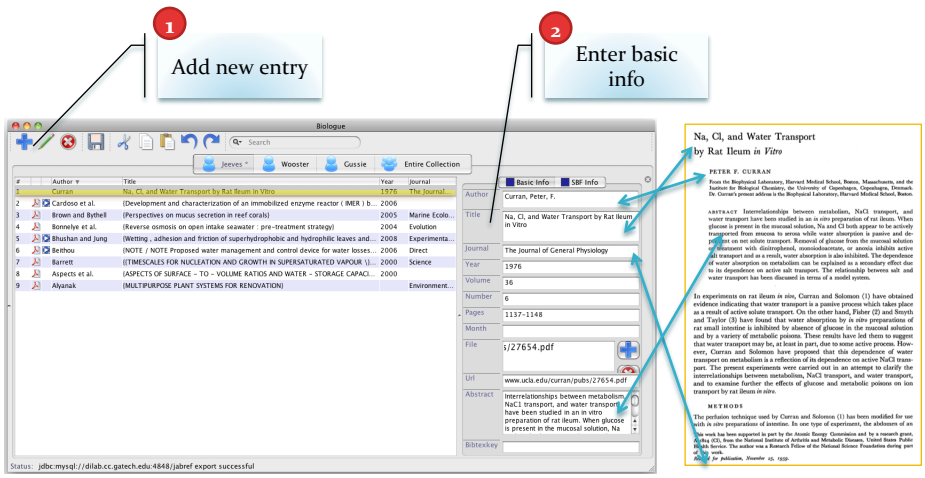


Figure 9.8: Posting a new citation in Biologue

Implementation of model-based tagging in Biologue

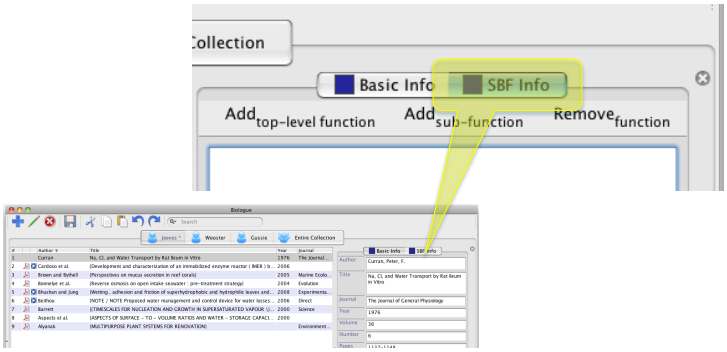
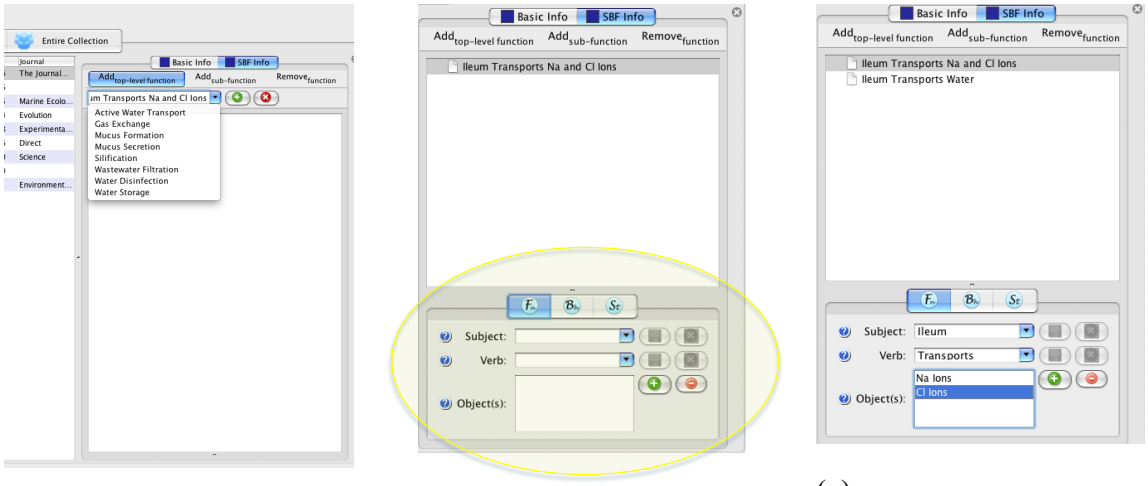


Figure 9.9: A separate tab for model-based tagging

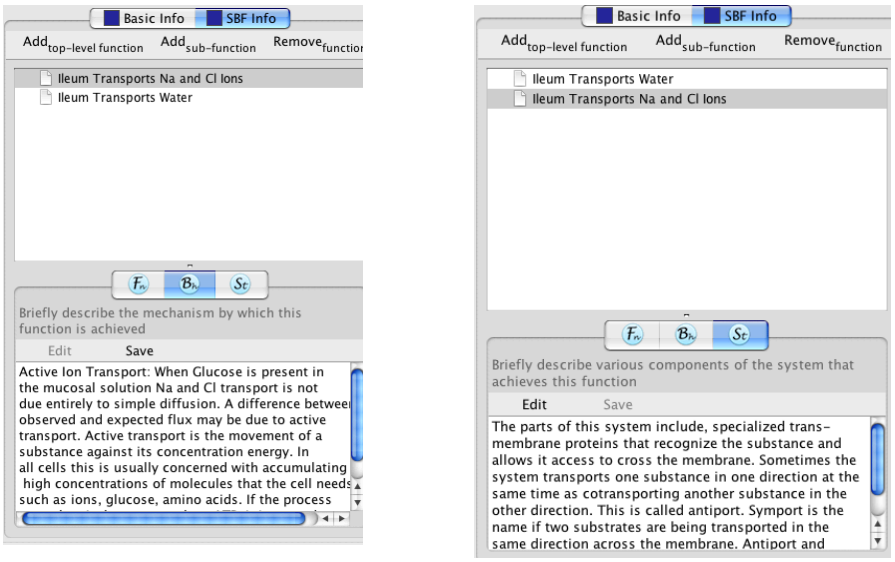
Once an article is posted in Biologue, users also have the option of tagging it with PSSBF models. This is done by clicking on a separate "SBF info" tab as shown in Figure 9.9. Once they are in the tagging mode, they can proceed to add a function of the biological system discussed in the article as shown in Figure 9.10(a). Note that Biologue provides auto-complete feature, so that when they start typing the name of the function, other functions in the Biologue repository that begins with the same name are suggested

to the user. This encourages reuse and also semantically linking the current article to other articles that have the same function.



(a) (b) (c)

Figure 9.10: (a) Tagging a function in Biologue; (b) additional details associated with the function



(a) (b)

Figure 9.11: Adding (a) behavior and (b) structure information pertaining to a particular function

Once a user tags the article with a function name, consistent with the PSSBF ontology, they see three additional tabs at the bottom: *Fn* (for function information), *Bh* (for behavior), and *St* (for structure), as shown in Figure 9.10(b). In the *Fn* tab, a user can add the subject, verb, objects, and operating environment information, as per the PSSBF ontology. This is shown in Figure 9.10(c). In the *Bh* tab, a user can add text that describes the behavior of the system that achieves the said function. This is captured in the snapshot shown in Figure 9.11(a). A user can also add physical principles associated with that behavior as per the PSSBF ontology. Similarly, users can add textual descriptions of the structure of the system particular to that function as shown in Figure 9.11(b).

Group management

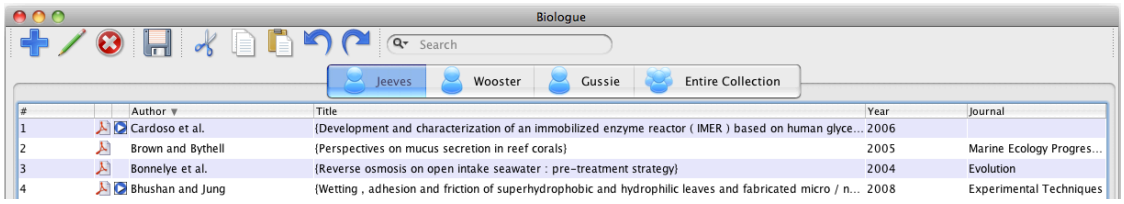


Figure 9.12: Groups in Biologue

A design team can form a group within Biologue. Groups are collections of users who can share their workspaces. They are useful for keeping track of what everyone else in the group is adding and reading. Groups are currently private in Biologue. The group feature was added because as we have seen, online bio-inspiration seeking is a collaborative activity. A feature that allows seamless sharing and monitoring of each other's information-seeking activity in the group is not only consistent with but also recommended by the kinematics of online bio-inspiration seeking (see Chapter 6 for more details).

Group membership is through invitation. Once a user joins a group, upon logging in, this user not only views his/her own workspace and the public collection of articles,

but also workspaces of other member in the group (shown in separate tabs in Figure 9.12). However, workspaces of others can be accessed in read only mode – they can view the articles in them, but cannot add, modify, or delete articles from other's workspace. Currently, there are no mechanisms that help users monitor recent activity in other users' workspace. They have to manually browse and keep track of what is new and what is not. However, it is easy to add facilities like activity RSS feed such that everything that appears under "Recent group activity" can be picked up by users.

Search and retrieval

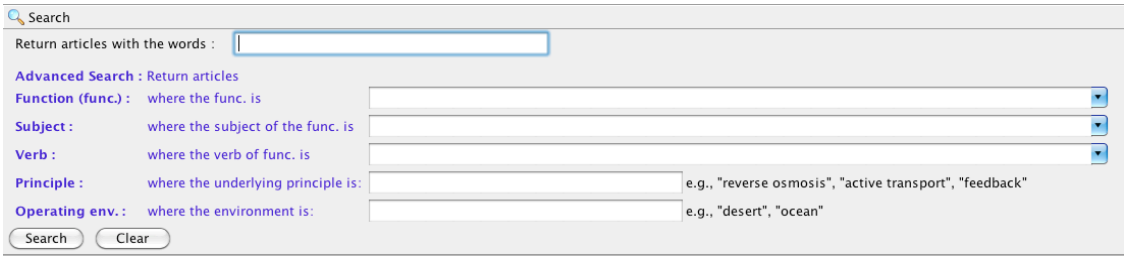


Figure 9.13: Advanced search feature in Biologue

By allowing users to annotate articles with SBF models, Biologue provides an advanced search facility that includes searching based on features like function, physical principle, operating environment, etc. that are derived from the SBF ontology. This is depicted in Figure 9.13. Because these features match the kinds of features that designers naturally tend to use when they are seeking bio-inspiration, it is expected that this search mechanism can address the problem of keeping the between-patch foraging cost low by making the information retrieval more targeted or focused to the needs of biologically inspired designers. A keyword-based search facility is also available in Biologue in case the advanced search facility does not provide good results.

Furthermore, the search mechanism in Biologue returns a list of relevant articles with a functional-decomposition overview of PSSBF models as part of the articles'

proximal cues. This is depicted in Figure 9.14. Again, it is our hypothesis that the inclusion of this overview can help information seeker make better judgment about which articles to attend to more closely and which ones to ignore. These overviews are interactive. Users can click on any function in this overview and get a more detailed model, including other information like behavior, structure, etc. associated with that function.

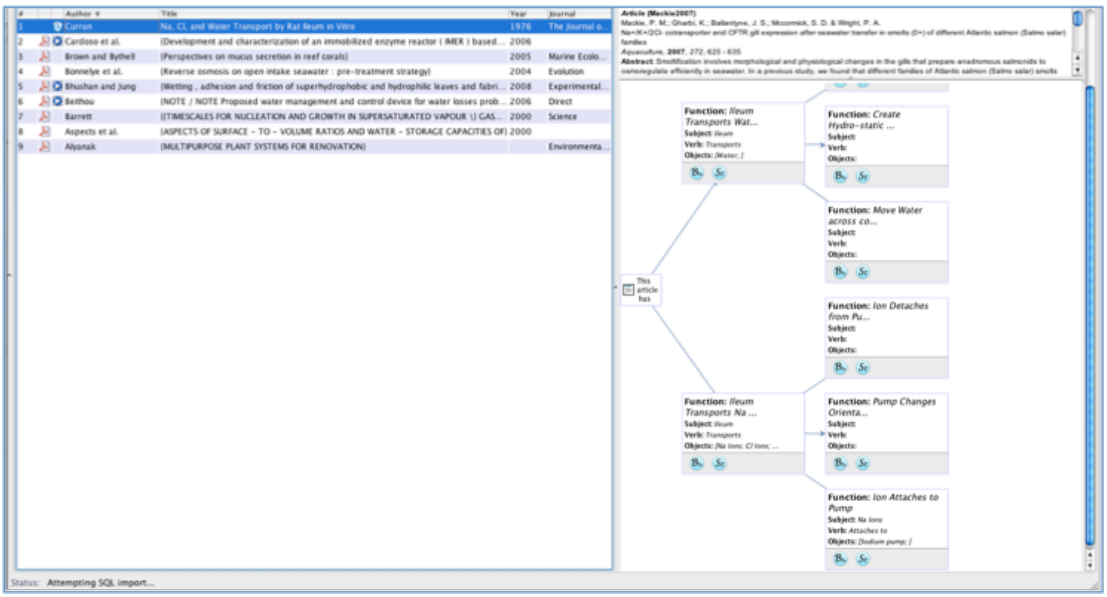


Figure 9.14: Search results

When a user double-clicks and opens an article in Biologue, it is displayed side-by-side to the interactive overview of PSSBF model associated with that article (assuming that this article has been tagged with). At any point during reading of the article, users can interact with and navigate the model, using it as a scaffold to understand the contents of the article.

Use-case scenarios

Let us first consider a resource-gathering scenario. This is a scenario where a biologist, in the course of her day-to-day work, comes across an interesting article on rat

ileum and how that organ passively transports water across osmotic gradient using a combination of forward- and reverse-osmosis. In such a scenario, Biologue allows this user to: 1) gather and organize this article in a personal library, 2) enter article's bibliographic information, 3) tag the article with an SBF model of the how the rat ileum works, and more importantly 4) share one or more articles (and models) with others.

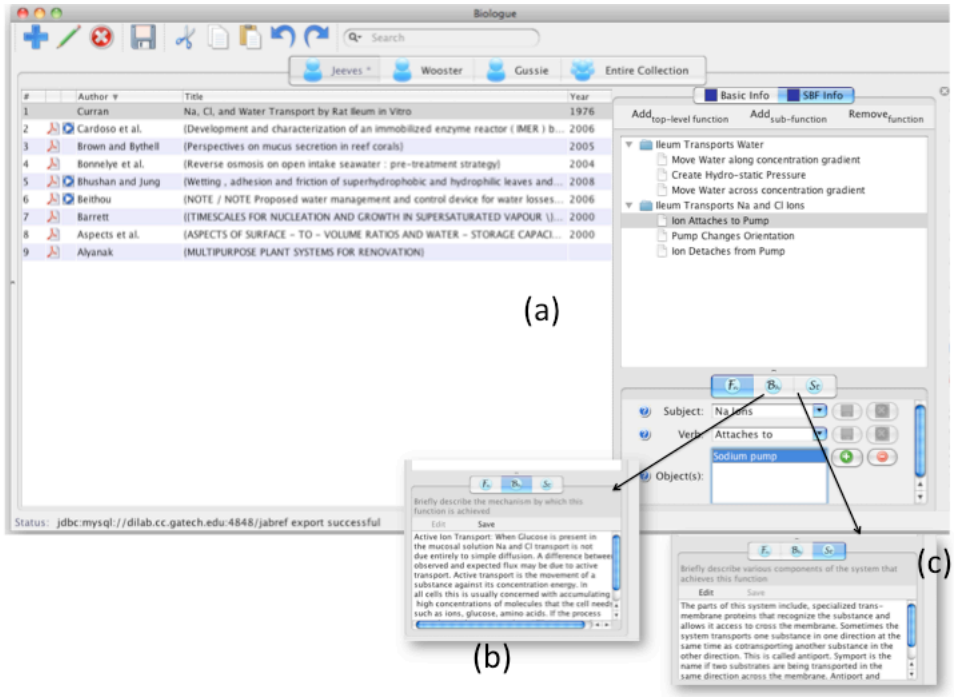


Figure 9.15: A newly added article in Biologue

Figure 9.15(a) shows the workspace of this user in Biologue with the newly added article depicted on the left hand side. On the right hand side, one can see a part of the SBF model that this user has created for this article. In particular, it shows the hierarchy of functions and sub-functions that are performed by this system. Associated with each function or a sub-function, the user has also added the behavior and structure information as shown in Figure 9.15(b) and Figure 9.15(c).

When a resource is tagged with a model, it is automatically indexed with facets derived from the model. For example, in the case of rat ileum article, it is tagged with

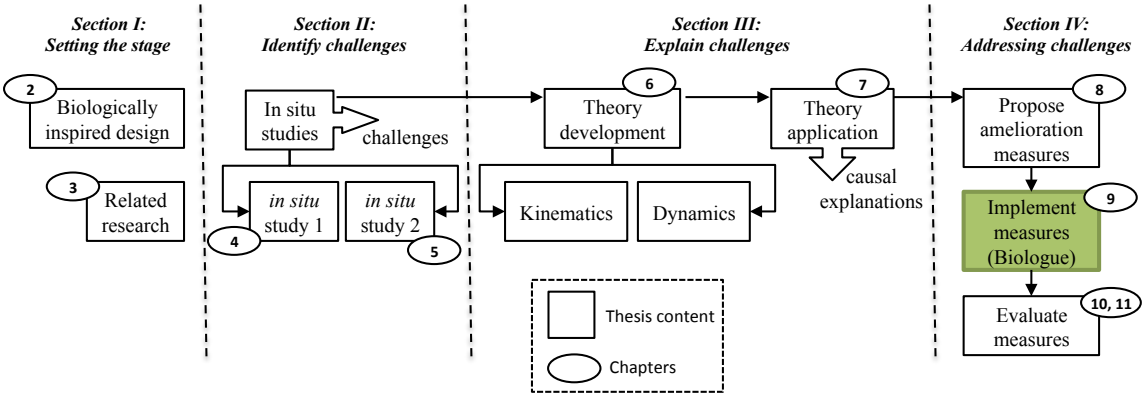
functions and sub-functions (e.g., ileum transports water, move water across concentration gradient), subjects of functions (e.g., ileum, serosa, mucosa), verbs of functions (e.g., transport, move, pump, attach, detach), physical principles (e.g., diffusion, osmosis, reverse osmosis), etc.

Now let us consider a resource seeking scenario: While the first task assists in the building of a corpus of knowledge that designers can draw upon, this second task is related to using that corpus to find biological sources of inspiration. Consider a second scenario where an engineer is trying to design a bio-inspired, energy-efficient, seawater desalination technique. This user logs in and proceeds to search the entire collection of articles in Biologue for a relevant biological source. There are two ways of searching articles in Biologue: traditional keyword-based search and advanced search as shown in Figure 9.13. Keyword search retrieves a non-ranked set of all articles in which the specified search terms appear, either as part of the title, abstract, or the body of the article. Advanced search allows retrieval based on features like functions, principles, operating environment, etc.

Let us assume that the engineer in this scenario chooses to search the collection of articles based on function “remove salt from water” and the principle “reverse osmosis”. This returns a non-ranked set of all articles whose SBF models are related to biological systems that satisfy both these conditions. When she clicks on the first article, the article about rat ileum, she can view not only the traditional bibliographic information that one would expect, but also the SBF model associated with the article. Studying this model gives her a gist of how the rat ileum works from an engineer’s perspective, which also helps her decide that it is worth pursuing the article. (The model also gives her a high-level conceptual framework that she can refer back and forth to guide her development of understanding of the article.) Upon reading the article and understanding how the rat ileum works in detail, she chooses to use this biological system as a source of inspiration

and develops a novel desalination technique that removes salt from water by a combination of forward- and reverse-osmosis.

Chapter Summary



This chapter discusses the design guidelines, architecture, system implementation, and use case scenarios of Biologue. The specification of PSSBF ontology, a functional modeling scheme derived from SBF, is also discussed in this chapter.

Biologue represents a social approach to establishing an online corpus of biology article references annotated by functional models. One unique feature of Biologue is *model-based tagging*, which represents a social approach to establishing an online corpus of biology article references annotated by their corresponding functional models. It is based on the principle of social bookmarking and is aimed to promote the sharing of biology articles among the designer community. In the same way that it is possible to catalog web pages (with Furl and del.icio.us) or photographs (with Flickr), designers can share information on biology articles in Biologue with semantic tools developed for the purpose of annotating and discovering articles using functional models. The functional models that are collectively annotated by the users are leveraged by Biologue to implement features such as model-based indexing and retrieval (first hypothesis), proximal cues enhanced by visual overviews (second hypothesis), and scaffolding for aiding designers' comprehension of biology articles (third hypothesis).

CHAPTER 10

EXPERIMENTAL STUDIES

In this chapter, I will begin evaluating the validity of the hypothesized measures for ameliorating the challenges of online bio-inspiration seeking (proposed in Chapter 8) using Biologue, which implements those measures (as discussed in Chapter 9). The research question pertinent to this chapter is:

RQ4: How does the presence of an online information environment with features grounded in the hypotheses associated with RQ3 change the dynamics of online bio-inspiration seeking in order to ameliorate the identified challenges?

In particular:

- ***RQ4.1: To what extent does changing the indexing and access mechanism from keyword-based to functional model-based impact the rate of encountering the relevant information resources?***
- ***RQ4.2: To what extent does including visual overviews derived from functional models in proximal cues impact the rate of recognition errors?***
- ***RQ4.3: To what extent does having functional models in addition to textual descriptions of biological systems impact designers' understanding of those systems?***

In this chapter, I will present three studies that I conducted in order to answer research question RQ4.1, RQ4.2., and RQ3. In the first *Find-frequency* study, we are interested in measuring the effect of function models-based indexing and access mechanism on the rate of encountering relevant information. In the second study, the *Error-Rate* study, we are interested in measuring the effect of function model overview enhanced proximal cues on the rate of recognition errors. In the third *Comprehension-*

study, we are interested in the measuring the effect of having functional models on the understanding of biological systems among novice designers.

Setup and procedure for Experiments 1 and 2

The participants for these studies were undergraduate Georgia Tech students who were compensated with gift cards for their participation. In the *Find-frequency* study, given a target bio-inspired design challenge, participants were assigned the task of finding as many relevant articles as they could within Biologue, and their performance was measured and compared under one of the two indexing and access schemes: functional models-based and keyword-based. In the *Error-Rate* study, given a different target bio-inspired design challenge, participants were assigned the task of rating the relevancy (on a scale of 1 to 5) of a small set of predetermined articles within Biologue based on articles' proximal cues alone, and their error rates were measured and compared under one of the two conditions: surrogates-with-SBF and surrogates-with-no-SBF. The length of each study was approximately 2 hours long.

The overall study procedure that was followed was:

1. Recruitment process: participant sign up for two two-hour sessions, one for each study
2. Session 1: *Error-Rate study*
    a. Pre-study/demographic questionnaire
    b. BID training (and take a test to ensure training was successful)
    c. Biologue training and testing (and take a test to ensure training was successful)
    d. Bio-inspired design challenge #1 (and take a test to ensure the challenge is well-understood)
    e. Perform the article rating task

3. A gap of at least one day for every participant between session 1 and session 2
4. Session 2: *Find-frequency study*
    a. Bio-inspired design challenge #2 (and take a test to ensure the challenge is well-understood)
    b. Perform information seeking task
5. Compensation

Participants

Participants were 16 students from Georgia Tech, a great majority of who were junior and senior undergraduate students. They were recruited by posting a message to the mailing lists of interest to BID community within Georgia Tech. All participants were offered gift cards as compensation for their participation. Both studies were between–subject studies, i.e., in each study, participants performed the assigned task only once and under one condition alone. Every participant signed on to two two-hour time slots that were convenient to their schedule, one for each study. For each study, participants were alternately assigned to one of the two groups on a first-come-first-serve basis. This ensured that (1) participants were randomly assigned to groups (because their assignment depended on their own schedule, over which researcher had no control), and (2) both groups were equally filled in both studies.

Study administration

Both studies involved the use of Biologue and therefore involved the use of a computer. The study was administered in the Design and Intelligence Lab at Georgia Tech. Participants had to arrive at the study location at the scheduled time. They were assigned a machine with Biologue preinstalled on it. If there were more than one participant taking the study at the same time, they were physically isolated from each other such that they were not visible to each other. Participants could walk out at any

time during the study and still collect their compensation, irrespective of whether they had completed the assigned task or not. However, to motivate them to complete the task and to perform it to the best of their ability, they were offered a performance bonus if they achieved a certain level of desired performance on the assigned tasks.

Researcher intervention during the study administration was minimal and was mostly restricted to proctoring activities like handing participants certain forms, getting their signatures, handing the task packets, etc. Participant' task packet was self-contained and included all the instructions required to complete the task with minimal interaction with the researcher. In fact, in many instances the study was proctored by people different from the researcher.

Experiment 1: Find-frequency study

In this study, we are interested in measuring the gain in rate of encountering relevant information resources as a result of employing functional models-based indexing and access mechanism. Given a target bio-inspired design challenge, participants were assigned the task of finding as many relevant articles as they could within Biologue, and their performance was measured and compared under one of the two indexing and access schemes: SBF-based and keyword-based.

Task description

The task that the participants had to accomplish in this study consisted of the following steps:

- Read and understand the problem statement describing the bio-inspired design challenge. They had to answer a test in order to ensure that they sufficiently understood the problem.
- Launch Biologue in search mode

- Try to find as many articles that are relevant to the given challenge as possible (upper limit of 14) in a stipulated amount of 2 hours
- During the course of the study, they were not allowed to search for articles beyond Biologue. They were only allowed to look up the meaning of technical terms if needed using specially designated topic reference websites.

Materials

The target problem: The bio-inspiration seeking challenge that was given to them in this study was called as *Ra Power*. This design challenge was related to solar thermal collector technology and involved the design of (1) a bio-inspired reflective panel that could be fitted onto a existing absorber, and was capable of dynamically changing its reflectivity, and (2) a bio-inspired feedback control system that regulates the temperature of glycol be regulating the reflexivity of the panel. This design challenge was not fabricated by the researchers, but was one of the actual challenges attempted by a design team (team Ra Power) in one of the earlier implementation of the BID course at Georgia Tech. The problem definition was adopted from that team's design report. The design challenge was presented to them as a presentation consisting of 4 slides as shown here:

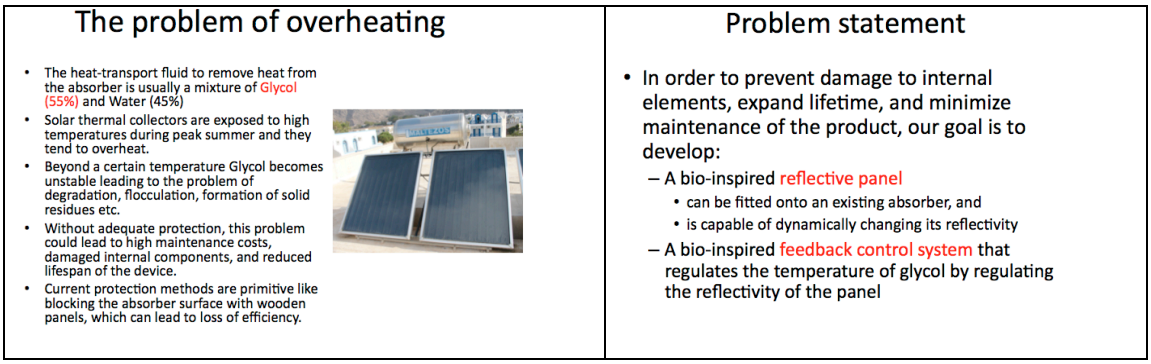
<table>
<tr>
<td>Slide 1

Solar thermal collectors

• Solar power is a free, reusable source for energy, which is very important in these times of climate change and dwindling resources
• Solar thermal collector technology harnesses this energy by absorbing sunlight and generating heat, which is then used to increase the temperature of the water flowing through the device, providing hot water for residential use.
  </td>
<td>Slide 2

Solar thermal collectors: Components

• A dark flat-plate absorber of solar energy
– Usually a sheet of thermally stable polymers, aluminum, steel or copper, often backed by a grid or coil of fluid tubing placed in an insulated casing with a glass or polycarbonate cover
• A transparent cover that allows solar energy to pass through but reduces heat losses
• A heat-transport fluid to remove heat from the absorber
– Usually a mixture of Glycol and Water
• A heat insulating backing. In water heat panels, fluid is usually circulated through tubing to transfer heat from the absorber to an insulated water tank.
  </td>
</tr>
<tr>
<td>Slide 3</td>
<td>Slide 4</td>
</tr>
</table>



Biologue: Two versions of Biologue were created for this study. In one version, the articles in Biologue's repository were indexed by keywords. Consequently, the articles were accessible only through keyword search. The search panel consisted of a single text box similar to Google and a search and a clear button. This version, called *Biologue-Vanilla*, was used in the one of the conditions of the study. In the second version, called *Biologue-SBF*, the articles in Biologue's repository were indexed by concepts and relationships that were part of the SBF models associated with those articles, including *functions*, *subject*, *verb*, and *objects* of those functions, *behaviors* and associated physical principles, and *structures*. Participants who used Biologue-SBF were able to search and retrieve articles based on all those features.

Biologue's repository had more than 200 articles in it. These articles were gradually collected through Biologue deployment in the BID course. These were articles posted by students of this course in earlier years. 14 special articles (referred to as *predetermined articles*) were included in the repository that were known to be relevant to the target design challenge. Team Ra Power cited these as articles they used in the course of their bio-inspiration seeking. If a participant's information seeking was very accurate, then nearly all these 14 articles would be found and reported by that participant and vice versa.

Reporting found articles: In the course of the task, once a participant found an article that they wanted to report as being relevant, they went to an online form and filled

one of the 14 slots in that form. In the slot for each article on the form, they could enter the title of the article and include a short description of why they thought that that article was relevant. Therefore, we not only collected the found articles, but also their rationale for choosing those particular articles.

Study design

This was a 1×2 between-subject design. It consisted of one independent variable and four dependent variables:

Variable	**Kind**	**Possible values**
Indexing and access mechanism	Independent	Keyword-based, SBF-based
Avg. find-period (inverse of find frequency)	Dependent	In the range 0 to 2 hours
Avg. number of information regions foraged	Dependent	In the range 0 to (some large number)
Avg. between-patch foraging time	Dependent	In the range 0 to 2 hours
Avg. yield per region	Dependent	In the range 0 to 1

Data

Participant demographic data was collected. The data for the study mainly came in the form of video data. The entire time spent by a participant on the computer during the study was recorded using screen-capture software. The other piece of the data was the found-articles data that was collected online for each participant.

Analysis

First, the participants in the two groups were compared to establish the equivalency of the two treatment groups. The following five features of participants were

used to establish this equivalency: 1) gender, 2) biology background, 3) design experience, 4) interdisciplinary research experience, and 5) the use of scholarly articles in their everyday work practices. There was no statistically significant difference between the two groups with respect to participants' gender (C^2 (1, N=16) =0.25, p=0.61), their biology background (C^2 (1, N=16) =0.29, p=0.59), extent of design experience (C^2 (1, N=16) =0.0, p=1.0), extent of interdisciplinary research experience (C^2 (1, N=16) =0.25, p=0.61), or extent of use of scholarly articles in their work practices (C^2 (1, N=16) =0.23, p=0.13).

Second, the found-article data was analyzed to determine: (1) the total number of articles found by a participant, (2) the number of predetermined articles within that total number. The rationale provided by the participant for selecting an article was analyzed to a lesser extent (just enough to make sure that nothing out of the ordinary was reported).

Third, the video data obtained for each participant was coded using the coding scheme provided in Appendix A. There were 16 videos to code, one for each participant. The origin of this coding scheme lie in the information foraging studies conducted by researchers in the human-information interaction community (Pirolli, 2007), which was developed to code and visualize the behavior of a person engaged in online information activity. The original codes naturally allowed depicting this behavior in the form of *web behavior graphs* (Card et. al., 2001). From these web-behavior graphs, collecting interesting statistics about the observed information seeking behavior becomes possible. A similar approach is adopted here. The video data analysis consisted of the following steps:

- A coding scheme was developed by the researcher based on the standard coding scheme for information foraging reported in the literature, but made minor changes in order to adapt that scheme to apply it to the present context.
- Two coders independently coded the 3 videos. We compared our codings and found that the inter-coder reliability was initially low (47.42%). We found that a

majority of the differences were in establishing the start and end times of an event. We ignored those differences where the timings were off by +/- 5 seconds. We also found that some of the differences in the codings were due to ambiguities in the interpretation of what those codes meant. We resolved those ambiguities by establishing standard norms for consistently coding such situations. After resolving those differences, the inter-coder reliability reached 89.11%. This meant that we could go on coding independently with a relatively high degree of agreement. We coded the rest of the 13 videos independently. The final inter-coder reliability, taking into account all the 16 videos, was 87.93%. In those approximately 12% of the cases where there was no consensus, the codlings of the experienced researcher was included.

- A small java program was created which took the coded data for each video as input, and obtained numerical values for various variables, including the four dependent variables mentioned above. This program also produced a visualization of the information foraging behavior in the coded data for each video. A total of 16 visualizations were produced. A sample visualization of the information foraging behavior of one of the subjects is shown in Figure 10.1.

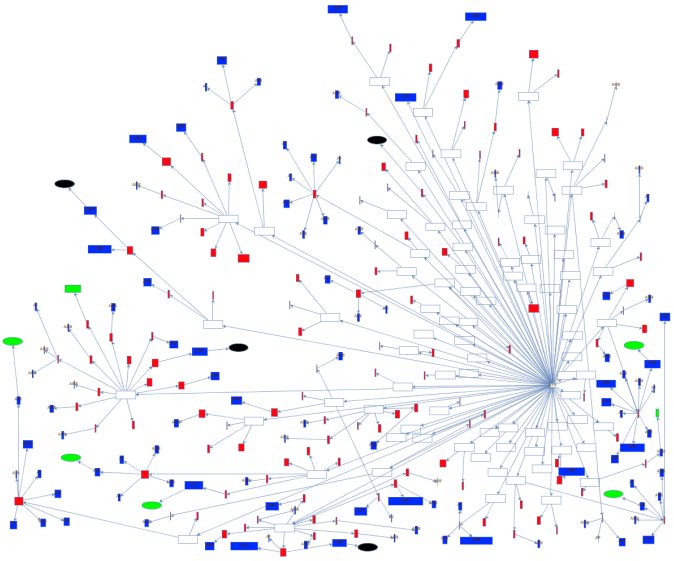


Figure 10.1: A sample visualization of information foraging behavior of a subject.

Expected results

I predicted that, in the experiment condition when compared to the control condition:

- *Find period* would be *less*
- *Between-patch time* would be *less*
- *Number of regions foraged* would be *less*
- *Yield per region* would be *less*

Actual results

We can look at each subject's episode as a whole and derive some performance characteristics, which is shown in Table 10.1

Table 10.1: *Episode-wide* average performance characteristics of the participants in the two groups.

Group		Total articles found	Total predetermined articles found	Total info. seeking time (mins)	Avg. find period (mins/article)
Biologue -Vanilla	avg.	8	2.63	83.63	11.48
Biologue -SBF	avg.	7.25	3.0	40.13	5.85

This suggests that there was *no* significant difference between the two groups with respect to the number of articles found. However, there was a significant difference between the two groups with respect to the cost incurred to find those articles. The Biologue-SBF group incurred less cost compared to the control group. Compared to Biologue-Vanilla, in Biologue-SBF:

- the average *information seeking time* was 52% less
- the average *find period* (mins/article) was 49% less

We can further look at each subject's period data (entire episode consists of many periods) and determine how and where their time and effort was spent in the process, which is shown in Table 10.2.

Table 10.2: *Period-wide* average performance characteristics of the participants in the two groups.

Group		Mean between-patch time (secs)	Mean within-patch time (secs)	Mean number of regions visited
Biologue-Vanilla	avg-mean	701	136.8	4.3
Biologue-SBF	avg-mean	177.81	141.17	2.45

This suggests that, compared to Biologue-Vanilla, in Biologue-SBF:

- the *average mean between-patch time* was 74.63% less
- the *average mean within patch time* was 3.1% more
- the *average mean number of regions visited* per period was 43% less

Finally, we can dig deeper into the data and look at number of information regions visited in each period and the information yield for each region, which is shown in Table 10.3.

Table 10.3: *Information region-wide* average performance characteristics of the participants.

Group		Mean yield per region
Biologue-Vanilla	avg-mean	0.07
Biologue-SBF	avg-mean	0.212

This shows that, compared to Biologue-Vanilla, in Biologue-SBF the average mean yield per region was 67% higher.

Discussion

The above results suggest that both the treatment groups were similar with respect to the quantity and quality of articles that they found during this task. But, the group that used Biologue-SBF significantly less time and effort compared to the group that used Biologue-Vanilla. In other words, for a similar output, the cost of information seeking in Biologue-SBF group was significantly less. If the performance of information seeking is measured as the ratio of total useful information resources obtained to the resource cost (time and interaction), then the performance of the group that used Biologue-SBF was much better compared to their Biologue-Vanilla counterparts.

Experiment 2: Recognition-Error study

In this study, we are interested in measuring the extent to which the amount of recognition errors change as a result of redesigning proximal cues to contain SBF information. Redesigning proximal cues such that designers are exposed to an overview of SBF-like models of biological systems of distal information resources during the between-patch foraging time reduces the recognition errors by improving the accuracy of information scent perception.

In this study, given a target bio-inspired design challenge, participants were assigned the task of rating the relevancy of a small set of articles in Biolgoue's repository based on the proximal cues of those articles. This set of articles contained an equal number of relevant and irrelevant articles. Their performance was compared under one of the two conditions: (1) proximal cues containing traditional elements like title, abstract,

and publication information, and (2) redesigned proximal cues to include SBF information in addition to the traditional elements.

Task description

The task that the participants had to accomplish in this study consisted of the following steps:

- Read and understand the problem statement describing the bio-inspired design challenge (this challenge was different from the one in the previous study). They had to answer a test in order to ensure that they sufficiently understood the problem.
- Launch Biologue in the recognition mode. In this mode, upon launching, every participant was presented with a list of article surrogates (not the entire articles).
- They were required to look at each of those article surrogates and rate the relevancy of those articles for the given target problem on a scale of 1 to 5.
- They were also required to provide a short rationale for their choice of rating.

Materials

The target problem: This design challenge involved the design of a bio-inspired desalination technique such that: (1) the salinity of output fresh water should be fit form human consumption, specifically drinking, and (2) the energy footprint of the new technique must be less than the existing industry-standard techniques. This design challenge was subject to the following simplifying assumptions: (1) the feed water is already filtered and pre-treated to remove all other unwanted contents, leaving designers to deal with only pure saline water, and (2) the design will not actively control for other parameters like pH and alkalinity, free residual chlorine, boron, etc. They were also given information about two existing industry-standard techniques for doing desalination, namely flash distillation method and reverse osmosis method. Some of the energy-related

problems associated with the industry-standard techniques were also presented. To sum up, they were given enough information so that a novice could be brought up to speed on the problem and had a rich enough mental model of the problem to be able to read an article and make a determination about its relevancy. They were also tested on their knowledge about this problem before they proceeded to perform the rating task. It is important to note that this was not a toy problem formulated by the researcher for the purposes of this study, but actually a problem taken from the field (problem that was attempted by team FORO in the BID course in Fall 2008).

The set of articles: A total of 8 biology articles were chosen for this task. These articles were not selected by the researcher *per se*. It was obtained from a pool of articles that the earlier team (team FORO) had researched in Fall 2008. 4 of those articles were confirmed as being relevant to solving the problem, and 4 of them were confirmed as being irrelevant for solving the problem. The judgment of whether an article was relevant or irrelevant was not the subjective decision of the researcher, but was obtained from the decisions made by team FORO with sufficient rationale.

Biologue: Biologue's repository for the purposes of this study consisted of only those 8 articles that were selected for the study. Biolgue for this study was instrumented such that as soon as a participant launched it, she would be instantly presented with a list of these 8 articles (surrogates only). This was meant to simulate a snapshot in the information seeking process where the seeker has just entered an information region (consisting of 8 articles) and needs to then prioritize the order in which these articles would visited based on the perceived relevance of each article to the target problem.

Two versions of Biologue were created for this study. In one version, *Biologue-Surrogates-Traditional*, participants saw the traditional version of surrogates, consisting of the articles' title, abstract, and publication information. In the second version, *Biologue-Surrogates-SBF*, participants saw the SBF-augmented version of the surrogates consisting of SBF model overviews, plus the traditional elements.

The SBF models of articles: To minimize research bias, the primary researcher recruited another researcher to build the SBF models of biological systems discussed in those articles. This model builder had not encountered the desalination problem and was not aware of the purpose to which the SBF models would be put to use. Therefore, he could not introduce bias by tailoring the SBF models to match the desalination problem. This process resulted in the creation of SBF models for all the 8 articles that was used in this study. These SBF models were then entered into Biolgoue and made available as part of the surrogates in one of the treatment groups.

Online relevance rating survey: The participants were required to rate the 8 articles whose surrogates were presented to them on a scale of 1 to 5. This was achieved by creating an online survey and asking the participants to take a survey when they were ready to rate the articles. The survey contained 8 questions, one for each article they were required to rate. The rating was couched as recommendation question: what their recommendation for the article would be for a team doing the desalination project on a scale of 1 to 5, where 1 represented "completely irrelevant (skip reading the article altogether)" and 5 represented "absolutely relevant (mimic the biological system in the paper and you will have solved the problem)." The middle value 3 represented "may be relevant, may not be relevant, can't say which." A portion of this survey for one question is depicted in Figure 10.2.

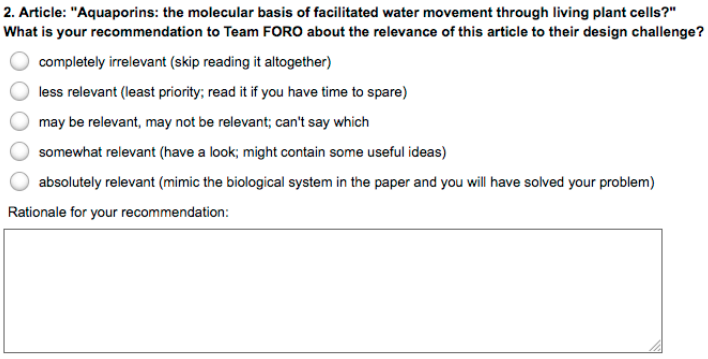


Figure 10.2: A sample question used to rate the relevance of an article in the RecognitionError study.

Study design

This was a 1×2 between-subject design. It consisted of one independent variable and four dependent variables:

	Variable		**Kind**	**Possible values**
	Surrogate design		Independent	Traditional, SBF-augmented
	Number of correct classifications		Dependent	0 to 8
Recognition error	Number of incorrect/undecided classifications	False positives	Dependent	0 to 8
		False negatives	Dependent	0 to 8
		Unable to classify	Dependent	0 to 8

The meaning of these variables in the context of this study is as follows:

- Correct classification: The article was actually relevant and the participant gave it a high score, or the article was not relevant and the participant gave it a low score.
- False positive: The article was not relevant, but the participant gave it a high score
- False negative: The article was actually relevant, but the participant gave it a low score
- Unable to classify: Irrespective of whether the article was actually relevant or not, the participant gave it a middle score (3, on a scale of 1 to 5)
- False positives, false negatives, and undecided's are considered as a failure to correctly classify an article, and hence are considered as recognition errors

Data

Participant demographic data was one of the data points used for this study. But the primary data for this study came from the online survey, which contained participants' article classification data, including the rationale for their classification.

Analysis

Although the participants were the same in both the studies, their distribution across the treatment groups was different in each study. Therefore, a group equivalency test had to be performed in this study as well. The same five features of participants were used to establish this equivalency: 1) gender, 2) biology background, 3) design experience, 4) interdisciplinary research experience, and 5) the use of scholarly articles in their everyday work practices. There was no statistically significant difference between the two groups with respect to participants' gender ($C2$ (1, N=16) =0.22, p=0.58), their biology background ($C2$ (1, N=16) =0.27, p=0.62), extent of design experience ($C2$ (1, N=16) =0.18, p=0.13), extent of interdisciplinary research experience ($C2$ (1, N=16) =0.29, p=0.44), or extent of use of scholarly articles in their work practices ($C2$ (1, N=16) =0.0, p=1.0).

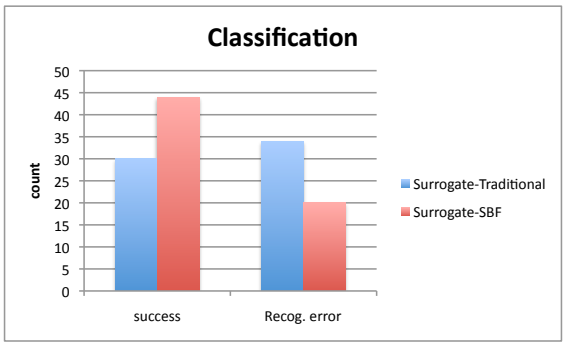
Second, participant classification data, which was on a 5-point scale, was converted into a 3-point scale. A value of 1 or 2 was classified as "irrelevant," a value of 4 or 5 was classified as "relevant," and a value of 3 was classified as "unclassified." For each participant and for each article, the participant classification was compared against the actual classification of the article. Based on this comparison, a determination was made as to whether it was a correct classification, a false positive, a false negative, or a no classification. A sample of this data for one participant is shown in Table 10.4.

Table 10.4: Sample article recognition data of one subject.

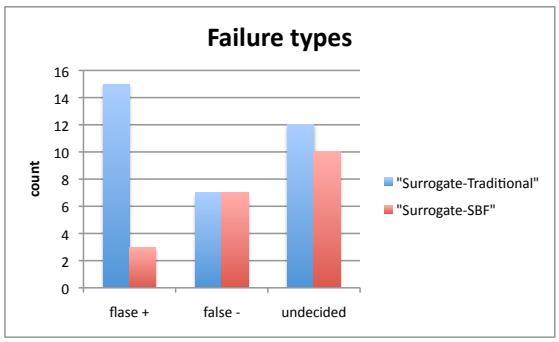
subID	treatment	article	correct	fPositive	fNegative	noClassification
350	0	1 (non-R)		1		
350	0	2 (non-R)				1
350	0	3 (R)	1			
350	0	4 (R)	1			
350	0	5 (R)				1
350	0	6 (R)	1			
350	0	7 (non-R)	1			
350	0	8 (non-R)				1
SUM			4	1	0	3
Total SUCCESS			4			
Total RECOGNITION ERRORS			4			

Results

A simple histogram analysis shows that the frequency of recognition errors in the Surrogate-SBF group was less than that of the Surrogate-Traditional group (see Figure 10.3a). A further breakdown shows that, to the extent where there is a difference in the frequency of recognition errors between the two groups, it is concentrated in false positives - there was no significant difference between the two groups with respect to false negatives and undecided (see Figure 10.3b).



(a)



(b)

Figure 10.3: (a) Classification histogram of two groups; (b) Histogram of failure types of two groups.

Table 10.5: Relevancy rating data of all the participants in the two groups.

subID	treatment	success	error	fPositive	fNegative	undecided
350	Surrogate-Traditional	4	4	1	0	3
300	Surrogate-Traditional	3	5	3	2	0
505	Surrogate-Traditional	3	5	2	1	2
29	Surrogate-Traditional	5	3	1	1	1
443	Surrogate-Traditional	2	6	3	1	2
635	Surrogate-Traditional	4	4	1	1	2
794	Surrogate-Traditional	4	4	2	1	1
748	Surrogate-Traditional	5	3	2	0	1
AVG		3.75	4.25	1.88	0.88	1.50
STDEV		1.04	1.04	0.83	0.64	0.93
225	Surrogate-SBF	5	3	1	0	2
720	Surrogate-SBF	5	3	1	1	1
154	Surrogate-SBF	6	2	0	0	2
863	Surrogate-SBF	7	1	0	1	0
420	Surrogate-SBF	6	2	0	2	0
279	Surrogate-SBF	7	1	0	1	0

512	Surrogate-SBF	5	3	0	1	2
641	Surrogate-SBF	3	5	1	1	3
AVG		5.50	2.50	0.38	0.88	1.25
STDEV		1.31	1.31	0.52	0.64	1.16

Table 10.5 shows that relevancy rating data of all the participants in the two groups. This data shows that, compared to Surrogate-Traditional, in Surrogate-SBF condition:

- the average *recognition error* was 41.8% less
- the average *false positives* was 79% less
- the average *false negatives* did not change
- the average undecided was 16.67% less

Discussion

The above analysis shows that in the context of this study, the group that worked with redesigned proximal cues containing functional model information, did significantly better with respect to the reduction in the average number of false positives. This reduction, for the most part, contributed towards the reduction in the average error rate that was seen in this group compared to the group that worked with traditional surrogates. It is not clear why there was no change in the average false negative or undecided rates. More fine-grained studies are required to determine the affordance of proximal cues vis-à-vis the different kinds of recognition errors.

EXPERIMENT 3

In a different but related line of research we, in our group, have been investigating the use of SBF modeling to enhance understanding of complex (ecological) systems in science education among middle-school students. In that context, empirical studies have

indicated that use of SBF models results in a deeper understanding of complex systems among middle-school students (Goel, Rugaber & Vattam, 2009; Hmelo-Silver et. al., 2008; Vattam et. al. 2010). The apparent success of SBF models as scaffolds for learning in middle school science inspired us to examine whether these SBF models may also lead to deeper understanding of complex biological systems among designers engaged in biologically inspired design. Here, I describe a study that investigates the effect of including SBF models on the understanding of complex biological systems among the current target research population, namely novice biologically inspired designers.

Study context and participants

This study was conducted as a classroom exercise for a group of 37 undergraduates enrolled in a biologically inspired design class at Georgia Institute of Technology. Of the 37 participating students, 16 self-identified as biologists and 21 as engineers. The participants were all junior and senior level undergraduates, fluent in English and familiarized with the concept of biologically inspired design through four weeks of classroom training.

This classroom exercise had both research and pedagogical goals. As a pedagogical device, the exercise served to (1) educate students on biological systems that might be useful to their design project, (2) familiarize students with differences in inferential capability afforded by different representation types, and (3) help students recognize patterns in communication and representation preferences among the different disciplines represented in the class. The pedagogical goals were realized both by participation in the exercise and by a reflective post-exercise discussion conducted after the exercise. The pedagogical goals served as additional incentive for the students to participate fully in the exercise.

One week prior to the exercise, the students received 90 minutes of classroom instruction in Structure-Behavior-Function (SBF) models. Aside from the pedagogical

benefits, this ensured that students were somewhat familiar with the SBF models presented during the study, although their fluency with SBF models probably did not approach their fluency with graphs or text. Furthermore, a five minute primer was provided to the students prior to the exercise, explaining the state representation schema for SBF models used in the SBF representations.

The cover page of each packet asked students to self-report on whether their major was biology or engineering, and how familiar they were with respect to the lotus leaf, the lotus effect, the basilisk lizard itself, and the basilisk lizard's water walking ability. Students were instructed to score their familiarity on a scale from 1 to 5, where one is totally unfamiliar, and five is very familiar.

Study Procedure

Students were provided one of three different modalities of detailed representations of two biological systems, and asked to answer questions about the systems along four dimensions:

- *fact finding*, the ability to find and return a single fact within the representation(s) provided.
- *spatial inference*, the ability to reason about or recall the shape or metric relationships among components described by the representation(s).
- *complex reasoning*, the ability to reason about casual and functional relationships among various components and interactions within the system described by the representation(s).
- *abstract problem solving*, *the* ability to answer complex questions related to the systems behaviors, but that were not explicitly present in the representation(s).

		Lotus Leaf			
		DIA	SBF	TXT	
Lizard	DIA	-	6	3	9
	SBF	5	-	6	11
	TXT	6	7	4	17
		11	13	13	

Table 10.6: Number of subjects by treatment type and model.

The treatments for each model were (1) text only, (2) text plus graphical and tabular representations, and (3) text plus structured representations. The structured representation was a Structure-Behavior-Function representation, discussed in detail in the next section. The students were given fifteen minutes to assimilate the new information and answer the questions, with a five minute period offered at the end for students who were not yet finished. The exercise was conducted twice, for two different biological systems, a lotus leaf and a basilisk lizard (lizard). These two systems were selected as representative of systems useful in the context of biologically inspired design. Each system was often cited by instructors in previous instances of the class, along with designs that were inspired by them. The Table 10.6 shows the combinations of treatments students received for the two different models. For the basilisk lizard, seven questions were asked: two fact finding, two spatial reasoning, two complex reasoning, and one abstract problem solving question. For the lotus leaf, five questions were asked: one fact finding, one spatial, two complex, and one abstract. Students that finished the first exercise early were instructed to close their packets, and not to look ahead to the second exercise. All students finished both exercises within the allotted time.

Exercise packets were pre-arranged such that a single student received two different modality combinations. Thus if a student had text-only modality for the Lizard, they would receive either text-plus-graphics or text-plus-structured-representation for the Lotus. This was important pedagogically so that students could reflect on differences in their own experience with the different modality combinations. This reflection was

facilitated by an instructor lead discussion following the exercise. Treatment types were alternated between adjacent participants, ensuring that roughly equal numbers of treatment types were distributed. Several non-student observers and instructors seated in the classroom also participated in the exercise. The results from these observers and instructors were discarded so as not to bias results. While it was our intent to test an equal number of each modality because of the distribution to observers and instructors, and subsequent discarding of their results, some imbalance occurred.

Furthermore, during the first round of exercises, some students did not look sufficiently ahead in their packets, and were unaware that they were given more than just the text representation. When students vocalized this fact at the end of the exercise, the test facilitators asked that any students who were unaware of the second, non-text representation during the exercise record this fact on their answer sheet. All answer sheets thus noted were considered text-only in terms of the analysis. This accounts for the disproportionately large number of text-only samples during the first exercise (17 of 37, versus 13 for the second). It also explains why 4 students received text-only versions for both models, as shown in Table 10.6.

At the end of the exercise, prior to the general discussion, on the last page of the packet students were asked to provide feedback on their preferred representation modality. The top of the piece of the paper read as follows: "In each case you were provided with different representations (either text with SBF, text with graphs/tables, or text only.) Which representations did you prefer? Why?" Students were allowed as much time as required to answer this question.

Materials used

This section discusses the materials used in this study, including the textual, SBF and imagistic representations of the two biological systems: self-cleaning function of lotus leaf, and Basilisk lizard's function of walking on water.

Text descriptions of the systems were extracted from papers describing the relevant details of their respective systems (Barthlott & Neinhuis, 1996; Hsieh & Lauder, 2004). The original papers were technical and difficult to read, and so were paraphrased to Flesch-Kincaid grade level score of 11.5. No mathematical formulae were present in the text descriptions.

Figure 10.4 illustrates the SBF model of the self-cleaning function of the lotus leaf, one of the biological systems used. The lotus leaf is interesting to engineers and others because it maintains a clean surface, despite being in otherwise dirty environments. It does this through nano-structures on the surface of the leaf that interact with water to cause it to bead up and roll off the leaf, carrying debris particles away with it. In Figure 10.4, *states* of a system are represented as shaded boxes, within which are described the *components* (e.g. contaminants, water droplets) and the *properties* (e.g. location, shape, mass) and *values* (e.g. on leaf, spherical, or the variable value M) associated with those components. For each state, we include only those components, properties and values relevant to the particular state change that is occurring. The entire series of state changes along with annotations about why the states change constitute the *behavior* of the system.

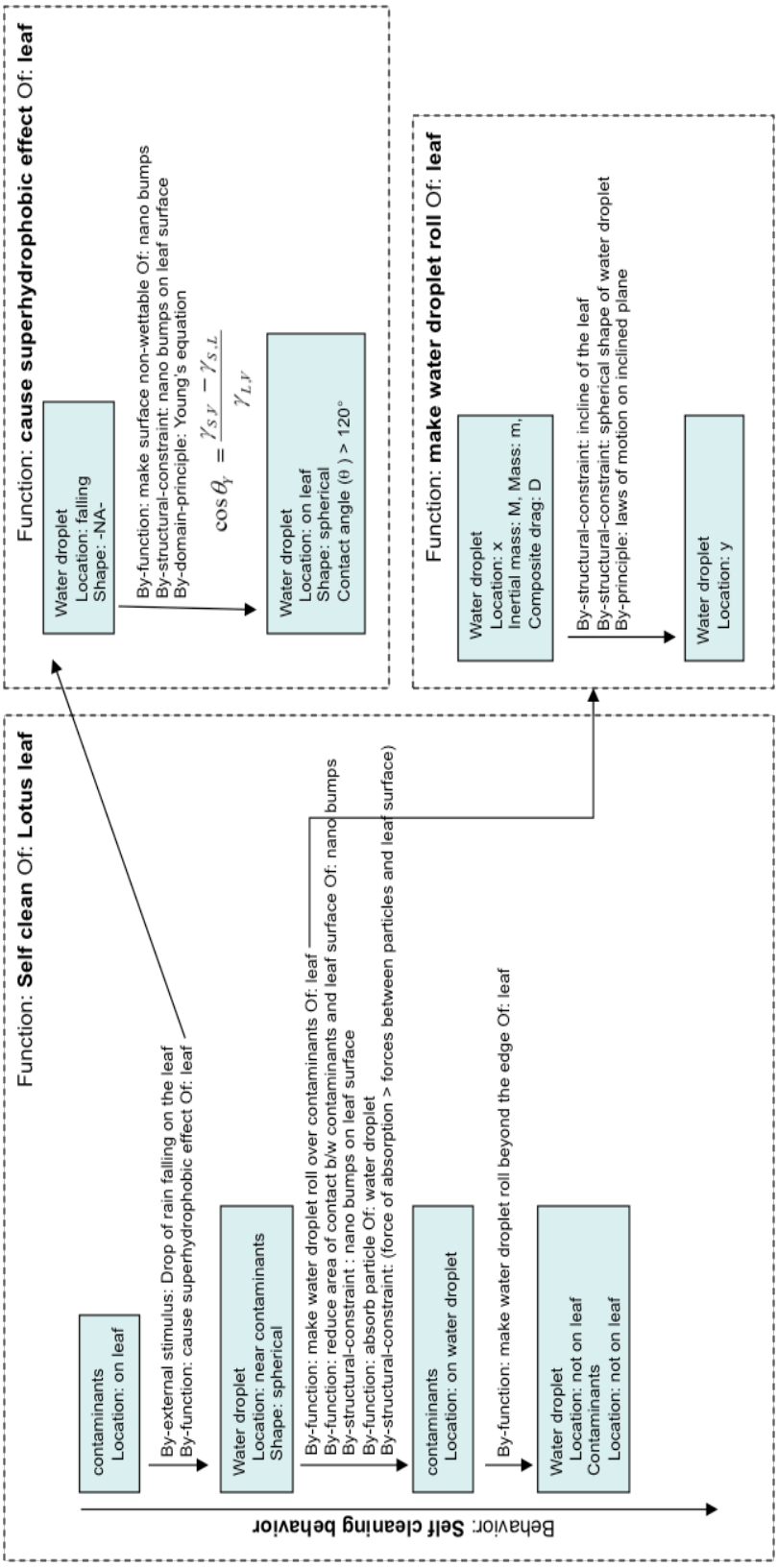


Figure 10.4: SBF model of self-cleaning function of the lotus leaf

Connections between states are called transitions, and include a variety of explanation types that provide information about why the change occurs. One type of transition, called transition-by-function, gives rise to the hierarchical organization of SBF models as we demonstrate in the following model.

The Self-Clean *function* (Figure 10.4a, on the left in Figure 10.4) of the lotus leaf is the result of a Self-Cleaning behavior consisting of four states. In the first state, contaminants are at rest on the lotus leaf. In the second state, when a drop of water falls on the surface of the leaf, the leaf exhibits a super-hydrophobic effect, which causes the water droplet to take the shape of a sphere. Figure 10.4b (on the top right in Figure 10.4) illustrates the super-hydrophobic sub-function; note the by-function annotation on the first transition in the Self-Cleaning behavior. The arrow between the states is the *transition*, while the annotation is the *explanation*. These annotations provide causal explanations for why the state changes occur in the system. The by-function annotation includes a pointer to a function that is represented by another SBF model, albeit a very small one. In this way SBF models inherently provide function /sub-function decomposition. In the third state, after the water drop falls on the surface of the leaf, the drop rolls over the contaminants using the principle of motion of a spherical body on an inclined plane, subject to the structural constraint that the leaf is inclined and not horizontal. Figure 10.4c (on bottom right of Figure 10.4) illustrates this sub-function; again, note the by-function *explanation* of the *transition* in the Self-Cleaning Behavior serves as the pointer to this sub-function, which itself is represented with an SBF model. In the fourth state, the drop of water rolls off the leaf, carrying the contaminants with them and leaving the leaf clean.

The Cause Superhydrophobic Effect sub-function of the leaf (illustrated in Figure 10.4b) has is its associated behavior which is enabled by the nano-scale “bumps” structural constraint present on the surface of the leaf, by the principle of interacting surface tensions captured by Young’s equation, and by the sub-function (not detailed in

this model) of the nano-bumps of making the surface non-wettable. The Make Water Droplet Roll function of the leaf (illustrated in Figure 10.4b) too has its own causal behavior. When the water moves over the contaminants, it absorbs them subject to the constraint that the force of absorption is greater than the static forces between the contaminants and the surface of the leaf. Note that the SBF model enables access to the physical laws and mathematical equations.

The lotus leaf model presented in Figure 10.4 provides a representative example of an SBF model. However, this characterizes only one way of visualizing an SBF model. For the second model used in our study, we modified the representation to better express states and transitions occurring in parallel. Figure 10.5 shows the behavior model of the basilisk lizard, which is interesting for its ability to quickly walk on water using only its hind legs. The state of the lizard, and the state of the water over which it is walking are represented on the left and right hand sides, respectively, with a common set of causal transitions in between. In this case, the sub-functions for the by-function explanations (e.g. Leg Slap, Push Water Down and Away, Exert Lift etc.) are not further modeled. The model itself captures only the essential functions and interactions useful for explaining how the basilisk lizard walks on water.

It is important to recognize that these models are qualitative. They do not seek to provide precise, mathematical models of a system per se, but rather to capture a conceptual understanding of how a system works. Because of their flexibility, it is not uncommon to see many differences between models developed independently by two individuals. We used SBF representations that explicitly captured the relationships between states, state properties, and the relationships between states (see Figures 10.4 and 10.5). The SBF models used were prepared earlier by the authors as sample SBF models for demonstration purposes. Graphical annotations present in these original SBF models were removed, and some formatting was altered for readability. All other content of the SBF models were preserved.

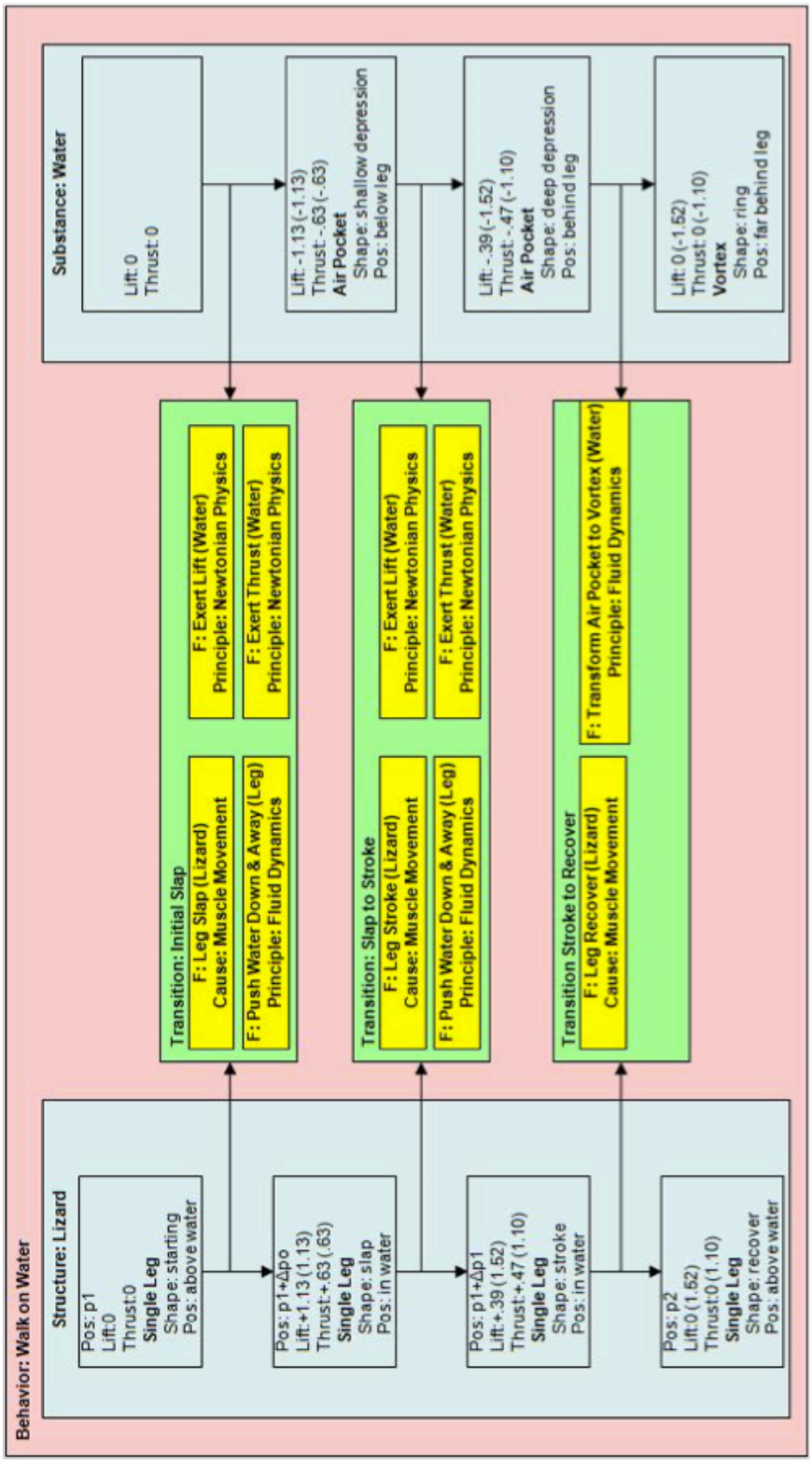


Figure 10.5: SBF model of Basilisk lizard walking on water

Finally, Figure 10.6 shows the graphical representations of the lotus leaf system including images of the systems (Figure 10.6a), and figures representing the operation of the system over a series of time ordered states (Figure 10.6b). Graphic representations were taken either directly from the corresponding academic papers, or from diagrams developed in our lab for use in augmenting SBF model descriptions, and were used without modification.

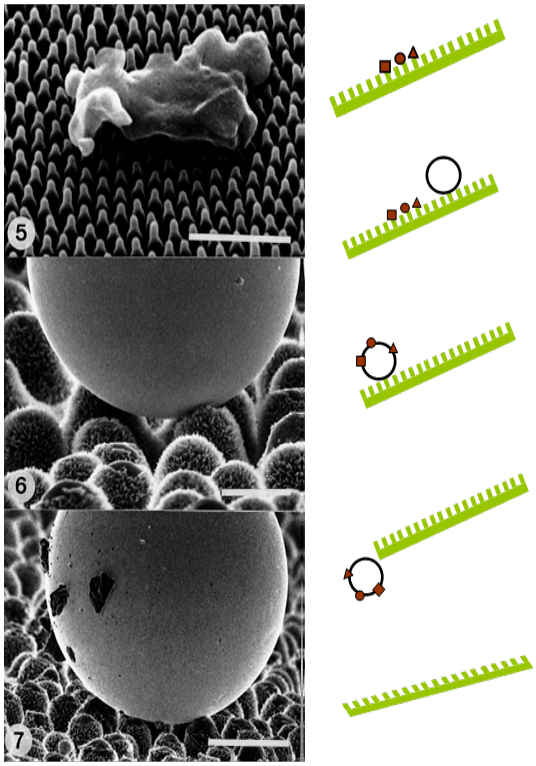


Figure 10.6: Diagrammatic representation of lotus leaf.

Each student was asked the same set of questions for each system. Following are a list of sample questions for both the (a) basilisk lizard system and the (b) lotus leaf system:

Fact finding:

(a) Which provides more lift, the slap phase or the stroke phase of the basilisk lizard's movement?

(b) What physical properties of the lotus leaf account for it being clean?

Spatial Inference:

(a) In which phase, slap or stroke, does the moving leg cover a greater total distance?

(b) What shape does the water droplet form on the leaf of a lotus leaf?

Complex:

(a) Which provides more thrust, the slap phase or the stroke phase of the basilisk lizard? Why?

(b) How does the water droplet move on the lotus leaf?

Abstract Inference:

(a) How could you estimate the thrust and lift generated by the basilisk lizard, without measuring anything about the lizard itself?

(b) How is this different from how water might move over a surface without the properties of the lotus leaf?

Grading method

As an informal study, answers to questions were graded by only one of the authors, a computer scientist, with neither biology nor engineering training. His knowledge of both the lotus leaf and basilisk lizard systems is derived from scientific research articles, developing SBF models of the systems, observing the biologically inspired design class, and from discussions with biology and engineering instructors in the class.

The correct answers to fact finding and spatial inference questions were unambiguous. The answers to complex questions, and abstract inference questions were subject to some interpretation, as discussed in the following section.

Data

Reported Score	Lotus Leaf	Basilisk Lizard
1	22	37
2	23	22
3	15	12
4	9	3
5	5	0

Table 10.7: Self-reported familiarity score by model

The self-reported familiarity scores are presented in Table 10.7. The self-reported mean familiarity for the basilisk lizard system was 1.74, for the lotus leaf, 2.35.

Answers to questions were categorized as either correct or incorrect. For complex and abstract questions, some unanticipated answers were received that were not initially classified as correct or incorrect, because of some ambiguity in the question language. For instance, when asking how the lotus effect is accomplished, a student might cite the underlying property accounting for the behavior (for instance hydrophobicity), or might describe the motion of the drop of water as it rolls down the leaf and pick up particles. Both are legitimate correct answers to the question. For such questions, any rational answer citing facts and following a logical thought progression were coded as correct. Where multiple correct answers were thus possible, which correct answer was provided was noted. For instance, when asked how a drop of water might proceed down a lotus leaf, the terms “rolls” “fast” “by adhesion” and “non-wetting” were all coded as rational and correct, and each given a unique identifier. For the purposes of this study, however, only the correctness of each answer was analyzed. Only obviously wrong answers were coded as wrong. For example, for the complex question “How does the water droplet move on the lotus leaf?” the answer “by spreading” was considered incorrect because it is

the opposite of the correct answer (the water maintains a spherical shape and specifically does not spread.) Non-answers (blanks), accounted for 4.7% of the total answers, and were provided a unique code but were considered incorrect for purposes of the analysis.

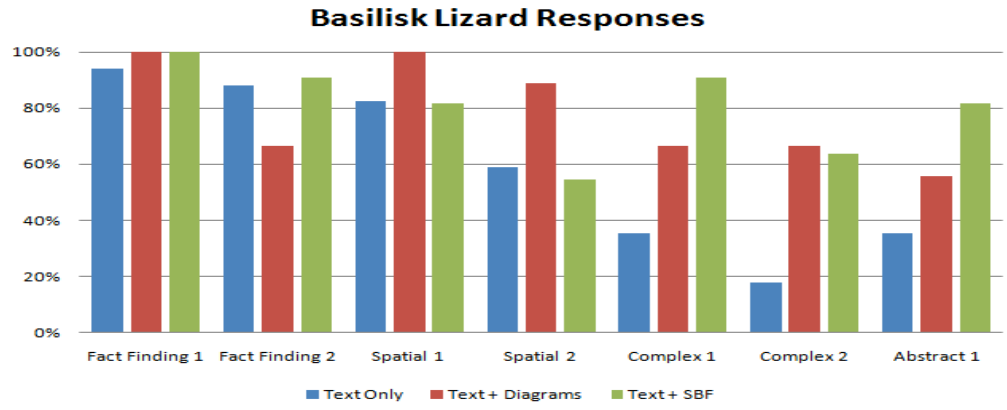


Figure 10.7: Percentage of correct response to Basilisk lizard question by treatment type

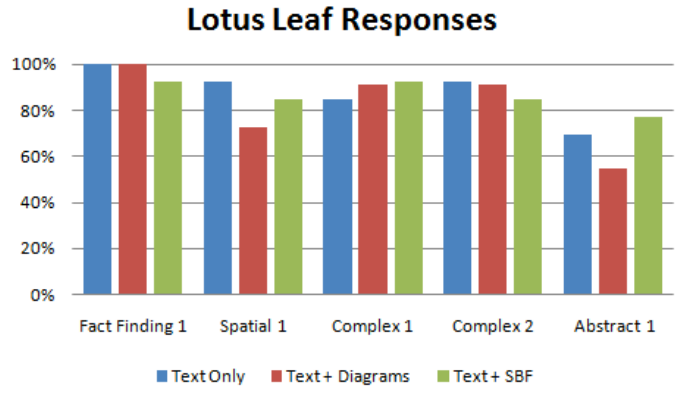


Figure 10.8: Percentage of correct response to the lotus leaf question by treatment type

For the basilisk lizard based questions, Figure 10.7 shows the percentage of correct answers for each question, by treatment type; Figure 10.8 provides the same information for the lotus leaf based questions. Table 10.8 reports the average percentage correct, by question, by major, irrespective of treatment.

With respect to the final question, preferred representation, interestingly some students felt strongly enough to not only comment on their preferences, but also to

comment on their dislike for the SBF modality. Table 10.9 summarizes student preference by major, where the row heading Not SBF represents the number of students that reported a dislike for the SBF modality.

Table 10.8: Percentage of correct answers by major

	Fact Finding 1	Fact Finding 2	Spatial 1	Spatial 2	Complex 1	Complex 2	Abstract 1
	Basilisk Lizard Questions						
Biologists	100.00%	87.50%	81.25%	56.25%	50.00%	31.25%	43.75%
Engineers	95.24%	80.95%	90.48%	71.43%	66.67%	52.38%	61.90%
	Lotus Leaf Qustions						
Biologists	100.00%		81.25%		93.75%	81.25%	43.75%
Engineers	95.24%		85.71%		85.71%	95.24%	85.71%

Table 10.9: Representation preference by major

Preference	Biologist	Engineer
Diagram	5	18
SBF	6	1
Text	5	2
Not SBF	1	4

Analysis

Familiarity Scores

Although the mean reported familiarity with lotus was greater than that for the basilisk lizard, and scores were generally higher for the lotus questions than for the basilisk lizard questions, correlation analysis between the self-reported understanding of a system and the number of correct answers show close to zero correlation (r-squared = .015 for basilisk, r-squared = .047 for lotus). Thus, self-reported prior knowledge of a system does not appear to be an important factor for this study. This is likely a result of the level of detail of the questions being asked relative to a student's perception of their own familiarity. While a student might be familiar with the basilisk lizard and the

function it performs as reported through popular media, for instance, it seems unlikely that they would know or retain the particular thrust ratios discussed in an academic paper.

Question Scores

The mean score for the basilisk lizard was 4.27 out of 7 (61%), with a standard deviation of 0.87 (12.4%), while the mean score per student for the lotus leaf was a 3.7 out of 5 (74%), with a standard deviation of 0.66 (13.2%).

When assessing the significance of including SBF and diagrammatic modalities, we test the hypothesis that the proportion of questions answered with SBF or diagrams is greater or less than the proportion answered for the base rate for text only for the same question, assuming standard normal distribution. We note that for the basilisk lizard questions, the number of students n = 17 for text only, n = 11 for text plus diagrams and n = 9 for text plus SBF. Diagram plus text results are statistically different at a confidence interval of .01 for *complex 2* (z = 2.68), and are statistically significant at a confidence interval of .10 for *spatial 1* (z = 1.34), *spatial 2* (z = 1.54), and *complex 1* (1.56). SBF + Text findings are significant at the .01 level for *complex 1*(z = 2.88), *complex 2*(z = 2.68) and *abstract 1*(z = 2.41) questions. For the lotus example, no significant differences were detected for any of the questions.

Likewise tests of significance between number of correct answers for each question were run between engineers and biologists. Statistically significant differences were detected between engineers and biologists for the complex 2 question for the basilisk lizard (z = 1.34) and for the abstract 1 question for the lotus (z = 2.55).

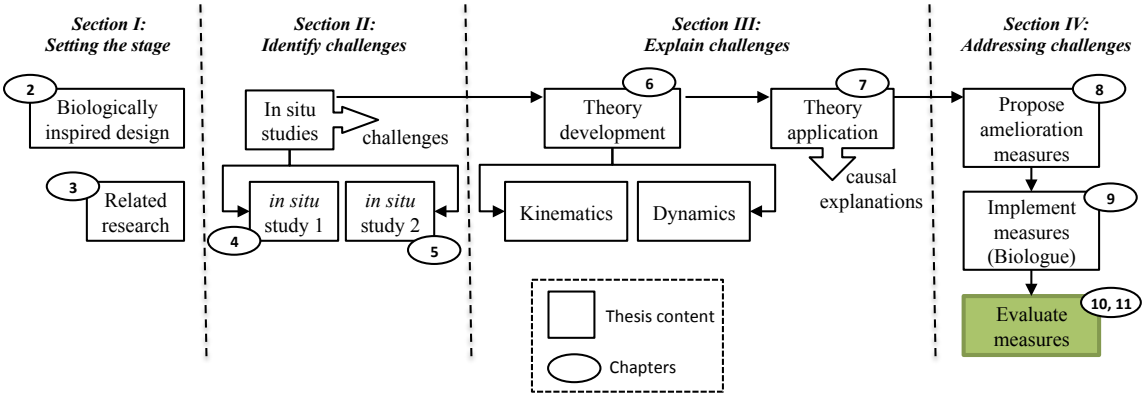
Table 10.10 summarizes the results, where High indicates statistically significant difference with 99% confidence, low indicates a statistically significant difference with 90% confidence. While not statistically significant overall, it is interesting and counterintuitive that for some questions, the additional graphical or functional information resulted in *worse* average performance. This can be seen in fact finding

question 2 for the basilisk model, and for spatial question 1, and abstract question 1 for the lotus leaf model.

Table 10.10: Summary of results

	Basilisk Lizard							Lotus Leaf				
	Fact Finding 1	Fact Finding 1	Spatial 1	Spatial 2	Complex 1	Complex 2	Abstract 1	Fact Finding 1	Spatial 1	Complex 1	Complex 2	Abstract
SBF + text	--	--	--	--	High	High	High	--	--	--	--	--
Diagrams + text	--	--	Low	Low	Low	High	--	--	--	--	--	--
Major	--	--	--	--	--	Low	--	--	--	--	--	High

Chapter Summary



This chapter presented three experimental studies in order to test the validity of the measures proposed in earlier for ameliorating the challenges of online bio-inspiration seeking. In the first *Find-frequency* study, we are interested in measuring the effect of function models-based indexing and access mechanism on the rate of encountering relevant information. In the second study, the *Error-Rate* study, we are interested in measuring the effect of function model overview enhanced proximal cues on the rate of recognition errors. In the third *Comprehension-study*, we are interested in the measuring the effect of having functional models on the understanding of biological systems among novice designers.

Trends from experimental study 1 indicate that the average rate of encountering relevant information resources was significantly higher when biology documents were indexed and accessed using functional models when compared to keyword-based indexing and retrieval. Trends from experimental study 2 indicate that the average number of recognition errors was significantly lower when proximal cues were enhanced with visual overviews derived from functional models when compared to conventional proximal cues. Trends from experimental study 3 indicate that the presence of functional models in addition to textual descriptions affords deeper understanding of biological systems when compared to textual descriptions alone. In the deployment study, overall trends in self-reported data indicate that having functional models in addition to biology articles helped subjects better recognize the relevance of articles, but did not help them during retrieval or comprehension. However, among a subset of subjects who used model-based tagging feature more extensively and contributed multiple functional models in Biologue, self-reported data indicate that having models did help them during comprehension, specifically, it made reading of articles easier and go faster. The trends from these studies are encouraging and seem to suggest that measure put forth indeed change the dynamics of online bio-inspiration seeking in favor of ameliorating the identified challenges of online bio-inspiration seeking.

CHAPTER 11

DEPLOYMENT STUDY

We introduced Biologue in the classroom context of ME/ISyE/MSE/PTFe/BIOL 4740 course in Fall 2010. In this term, the class consisted of 44 students. This was a project-based course where students were grouped into teams and each team was required to complete two biologically inspired design projects over the course of one semester. We introduced Biologue during a time when they were researching biological systems for their first project. Students were encouraged to (but were not required to) use Biologue to post citations of articles that they found relevant for addressing their design challenge and also to tag those articles with SBF models. Additionally, they were also encouraged to search for articles in Biologue.

We conducted a user survey at the end of the project to elicit students' impressions of Biologue after having used it for the duration of the project. In particular we wanted students to self-report on the usefulness of the advanced search feature of Biologue for retrieving relevant articles and the usefulness of SBF models for scaffolding their understanding of the articles they came across in Biologue. Out of the 44 students, 21 students responded to the survey request. The responses from these 21 survey participants are analyzed here.

Survey design

The survey design consisted of 15 questions. The following is a sample of some of the questions contained in the survey that are relevant to the analysis presented here.

2. How often did you use Biologue after it was introduced in the classroom?

- ○ Every assignment thereafter
- ○ Only for some assignments
- ○ Only for one assignment
- ○ Never

3. How often did you perform the following actions with Biologue?

	Frequently	Often	Sometimes	Rarely	Never
Adding an article to your workspace	○	○	○	○	○
Adding/editing bibliographic information to the article you added	○	○	○	○	○
Tagging the article you added with with SBF model	○	○	○	○	○
Searching the collection of articles with keywords	○	○	○	○	○
Browsing the collection of articles based on SBF model features (e.g., function, behavior, key principles, etc.)	○	○	○	○	○

4. How satisfied are you with the following features of Biologue?

	Very satisfied	Satisfied	Neutral	Dissatisfied	Very dissatisfied	Did not use this feature
Adding an article to your workspace	○	○	○	○	○	○
Adding/editing bibliographic information to the article you added	○	○	○	○	○	○
Tagging the article you added with with SBF model	○	○	○	○	○	○
Searching the collection of articles with keywords	○	○	○	○	○	○
Browsing the collection of articles based on SBF model features (e.g., function, behavior, key principles, etc.)	○	○	○	○	○	○

7. How much do you agree with the following statements about tagging articles with SBF models in Biologue?

	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
Having models associated with articles is advantageous for the reader because the models helps the reader quickly determine whether an article is useful or not	○	○	○	○	○
Having models associated with articles is advantageous for the reader because the models make understanding the articles easier	○	○	○	○	○
Having models associated with articles is advantageous for the reader because the models make the reading go faster	○	○	○	○	○
Constructing the models to tag my articles compelled me to understand the article more thoroughly	○	○	○	○	○
Learning to construct models in Biologue was easy	○	○	○	○	○
Once learnt, constructing models to tag the articles with was easy	○	○	○	○	○
I might consider tagging articles with models even outside of this course	○	○	○	○	○

Biologue survey results

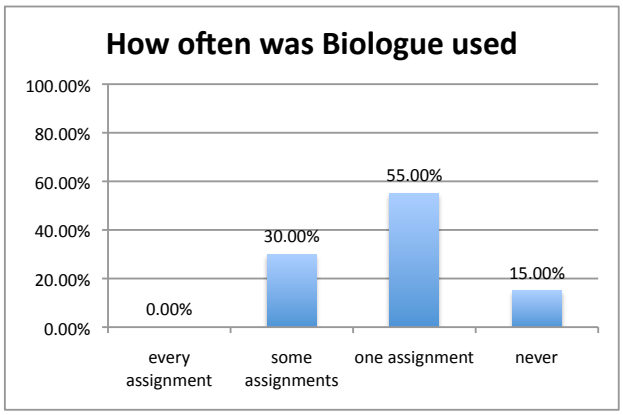


Figure 11.1: Overall Biologue usage

The usage of Biologue was not mandated in the class. Students were recommended to use Biologue as a research management tool for their assignments. As indicated in Figure 11.1, most of the students used Biologue for only one assignment, but some used it for more than one assignment. A small percentage of students also reported not using Biologue at all.

Resource gathering mode

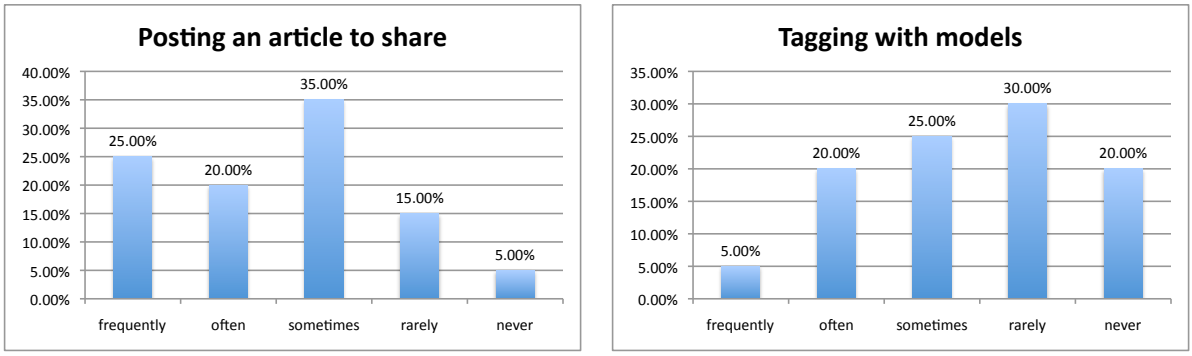


Figure 11.2: Biologue usage in the resource-gathering stage: (a) for posting articles, and (b) also tagging them with models

Among those students who used Biologue, a majority of them tended to post articles neither *frequently* nor *rarely*, but somewhere in between (on an average, 2 to 3 articles per assignment per student) (see Figure 11.2a). But a majority of them *rarely*

tagged the articles with models (Figure 11.2b). Among the models that were tagged, most of them were of depth 1 (containing one function and no sub-functions).

Resource seeking mode

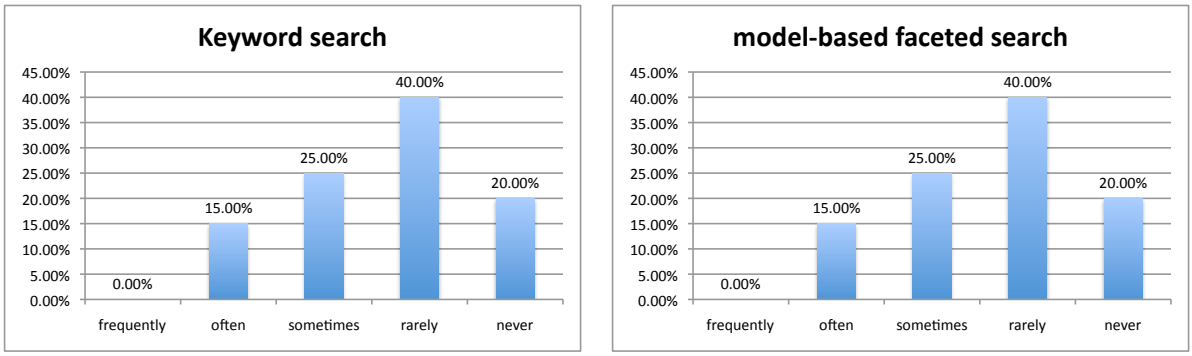


Figure 11.3: Biologue usage in the resource-gathering stage: (a) using keyword search, and (b) using model-based search

According to Figure 11.3a, students seldom engaged in searching for articles in Biologue using keyword-based search mechanism. Likewise, according to Figure 11.3b, students rarely engaged in model-based semantic search either. On the whole, the search feature was by and large neglected by the students. One possible reason for this is because of the limited size and coverage provided by the then existing Biologue repository of biology articles. The repository was indeed very small.

Usefulness of model-based tagging

Model-based tagging is one of the unique features of Biologue. We were interested in how this feature was received among the participants in the field. In particular were interested in the following questions:

- Does the presence of tagged SBF models provide an advantage in determining whether an article is useful or not?
- Does the presence of tagged SBF models provide an advantage for easier understanding of an article?

- Does the presence of tagged SBF models provide an advantage for faster understanding of an article?
- Does the act of tagging articles with SBF models compel users to gain a deeper understanding of the article?
- Did the users find the process of tagging articles with SBF models easy or complicated?
- Would users consider model-based tagging in other contexts outside the scope of this class?

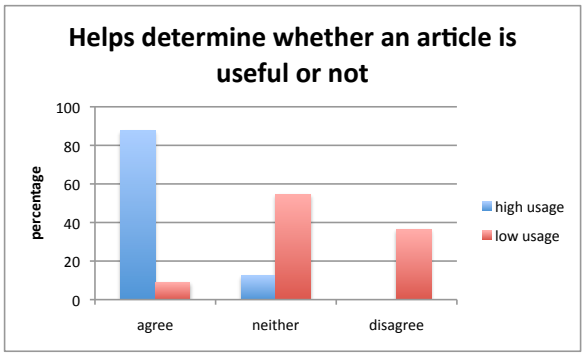


Figure 11.4: Usefulness of models for recognizing the relevance of an article

In response to the question related to the help offered by tagged SBF models to aid *recognition* (determine an article was useful or not), roughly 42% agreed that models helped, 21% did not think that models helped, and 37% could not decide.

We then classified students who responded to the survey into two groups: (1) high model-usage group, students who used the modeling feature to a greater extent (more than 75% of the articles they added were also tagged with SBF models), and (2) low model-usage group, students who did not used the modeling feature to a lesser extent (less than 75% of their articles were tagged with models). The more model-usage group consisted of 10 students (4 mechanical engineers, 2 industrial engineers, 2 biologists, and

2 others). The less model-usage group consisted of 11 students (3 mechanical engineers, 4 biologists, and 4 others).

As shown in Figure 11.4, comparing the two groups, there is a significant difference in the response to the same question regarding the help offered by SBF models to aid recognition. Among the high model-usage group, 87% agreed that the models helped, where as in the low model usage group, only 9% agreed that the models helped. So, users who tended to model more reported that the presence of models helps them make better decisions about relevancy. This indicates that those users who cared to use SBF modeling feature at storage time also derived the benefits of models at access time.

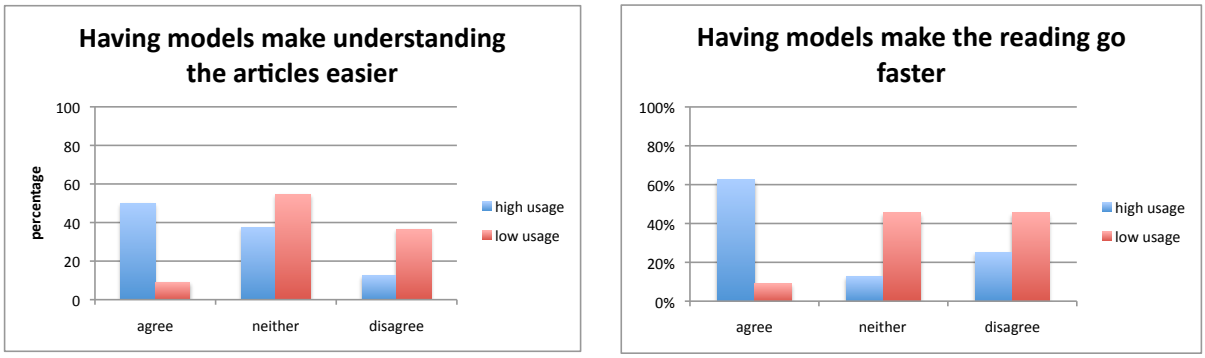


Figure 11.5: Usefulness of models for providing help in understanding articles: (a) models make understanding easier; (b) models make understanding go faster

Similar effects were found with respect to the question related to the help offered by tagged SBF models to aid *understanding* of biology articles (See Figure 11.5a and Figure 11.5b). Among the high model-usage group, 50% agreed that the models made understanding easier, where as in the low model usage group, only 9% agreed that the models made understanding easier.

Similarly, among the high model-usage group, 62.5% agreed that the models made understanding go faster, where as in the low model usage group, only 9% agreed that the models helped understanding go faster.

Other model-based tagging considerations

There are three other questions that I considered with respect to model-based tagging. First, whether the act of constructing models during tagging an article compels the tagger to deeply understand that article in the first place. If that is the case then there is a greater likelihood of model-based tagging leading to the emergence of richer and deeper models. Second, whether the task of tagging articles with models relatively easy or not. We want the cost of model-based tagging to be relative low for it to be useful. If the upfront cost of this task is high, then users are not likely to use that feature. Third, whether users are enthusiastic enough about the model-based tagging concept that they might use it other interdisciplinary contexts beside biologically inspired design.

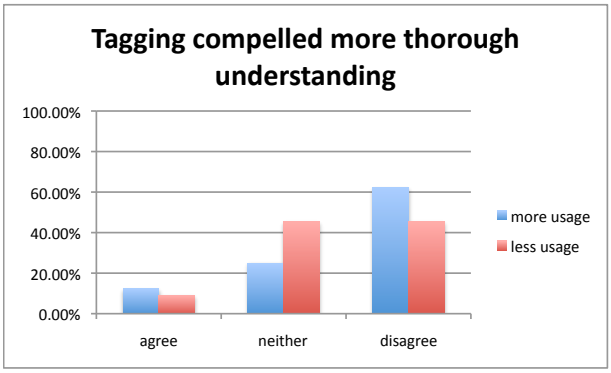


Figure 11.6: Model-based tagging necessitated deeper understanding of articles

Regarding the question of whether model-based tagging of articles necessitated a deeper understanding of articles in the first place, responders across both groups disagreed with that assessment (see Figure 11.6). This might explain why a majority of the models that were tagging in Biologue were shallow.

Regarding the question of whether model-based tagging was easy, among the high model-usage group, 75% agreed that model-based tagging was easy, where as in the low model usage group, only 8% agreed that it was easy (see Figure 11.7).

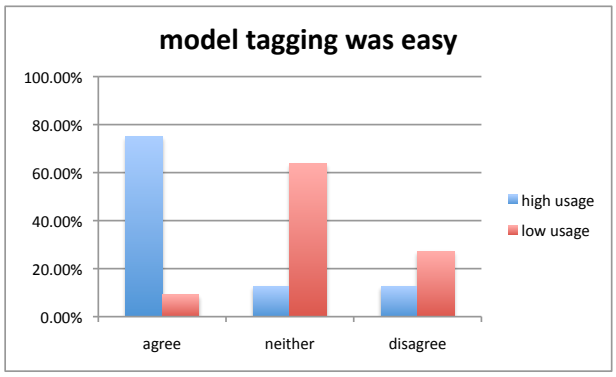


Figure 11.7: Model-based tagging was easy

Finally, on the whole, only a small number of participants agreed that they would consider tagging articles with models outside the scope of this course (10%). This opinion did not change even when we consider the two groups separately; among the high model-usage group, only 12% agreed that they would consider model-based tagging outside this course, and among the low model usage group, that number was 10% (see Figure 11.8).

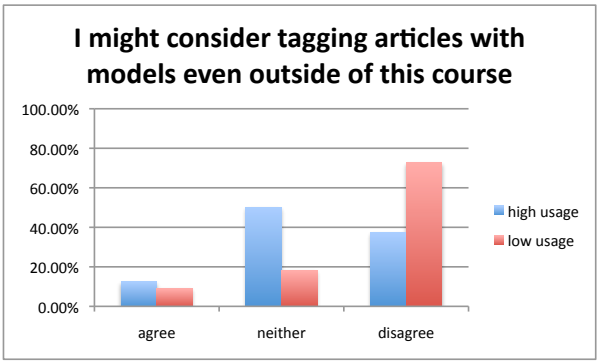
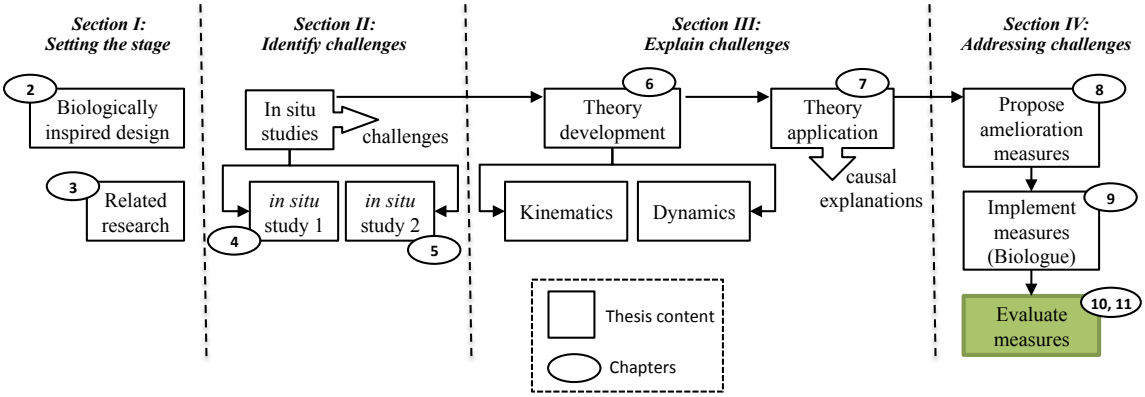


Figure 11.8: Model-based tagging outside the scope of the course

Discussion and Summary



This chapter presented a pilot deployment study of Biologue in the classroom context of ME/ISyE/MSE/PTFe/BIOL 4740 course conducted in Fall 2010. Students were encouraged to Biologue, but not mandated. A survey was taken at the end of their projects in order to find out their extent of usage of Biologue and how useful it was to them, especially the feature of model-based tagging. Results from survey suggest that users who cared to use SBF modeling feature at storage time also derived the benefits of models at access time. However, the fact that only a few students actually tagged their articles with SBF models indicates that the cost of creating SBF models might be high. At this stage we are cautiously optimistic about the potential of the model-based tagging to help designers construct understanding of biology provided we motivate them use this technique more often. But we do recognize that these numbers are merely suggestive and that we cannot make any strong claims based on this survey alone. More field testing and a larger sample size are required at this stage to make stronger claims.

CHAPTER 12

CONCLUSION

Biologically inspired design (BID) is a new design paradigm that has recently gained popularity because of its ability to generate innovative and sustainable technological advancements. Biologically inspired design uses analogies to biological systems to generate ideas for the conceptual phase of engineering design. The practice of BID aims to usher in a fundamental shift in engineering that is driven by biological sciences.

The optimism around BID is mostly based on important case studies of technological innovations that testify to the promise that this approach holds. However, a closer examination shows that there are important practical challenges associated with this practice. One of those challenges is finding the right source of biological inspiration given a target design challenge. This task of bio-inspiration seeking is intellectually challenging because of the vastness of the domain of biological systems and the relative lack of familiarity of the domain among designers coming from engineering.

The emergence of World Wide Web has made online information seeking a daily activity for most people. Whether at work, school, or play, people have come to expect instant access to information on any topic at any place and time. This expectation carries over to the task of seeking bio-inspiration as well. It is a common practice among novice designers to search online in order to find their biological sources if inspiration. However, due to their generic nature, current common online information environments on which designers rely upon do not provide adequate support specific to the particular circumstances of bio-inspiration seeking. Therefore, in spite of having online access to vast amounts of biological information, designers often struggle to find their biological sources of inspiration using the online approach. The reliance on online information

environments coupled with the lack of adequate support in those environments makes an intellectually challenging task even more difficult to achieve.

But in order to deal with what kind of specific support is needed, one needs to better understand the task of online bio-inspiration seeking and the challenges that designers currently face when engaged in this task. But at present, biologically inspired design research has not paid sufficient attention to understanding the nature of the bio-inspiration seeking task and its associated challenges. We know very little about the *in situ* practices of designers engaged in this task, nor do we sufficiently understand its underlying processes or 'mechanisms.' Yet, a majority of existing technology-building efforts in the domain of biologically inspired design research has focused on developing tools and techniques for aiding designers engaged in the task of bio-inspiration seeking. Therefore, there exists a gap between the research of biologically inspired design practice and the technology-building efforts for aiding the practice. Symptomatic of this gap, the current technology-building efforts tend to be technology-centric (as opposed to human-centric), whose design and development are craft-driven (as opposed to theory-driven).

This dissertation is an attempt to rectify this status quo in biologically inspired design research. Through a series of *in situ* studies, this dissertation uncovered the characteristics of and the challenges associated with the task of online bio-inspiration seeking. As a next step, grounded in those studies, a theoretical understanding of the process of online bio-inspiration seeking was constructed. The purpose of this theoretical account was to explain the challenges of online bio-inspiration seeking task and to identify the causes underlying those challenges. Once the causes were identified, a set of measures were put forth to ameliorate those challenges by targeting the identified causes. These measures were then implemented in an online information-seeking technology designed to specifically support the task of online bio-inspiration seeking. Finally, the validity of the proposed measures was investigated through a series of experimental studies and a field deployment study. This research is still in its early stages, but the

trends are encouraging, which seem to suggest that the proposed measures has the potential to change the dynamics of online bio-inspiration seeking in favor of ameliorating the identified challenges of online bio-inspiration seeking.

The Section II of this dissertation was concerned with identifying the challenges of online bio-inspiration seeking task. Based on the two *in situ* studies presented in Chapters 4 and 5, the following claim is made:

Claim 1: Designers engaged in online bio-inspiration seeking face at least three fundamental challenges. First, designers experience a *low rate of encountering relevant information resources* in online environments that they normally rely on. Second, designers experience a *high rate of recognition errors*: they fail to recognize the true relevancy of the information resources that they encounter in those information environments. Third, designers experience *significant difficulty in comprehending information resources* that they recognized as being relevant and struggle to develop conceptual understanding of biological systems discussed therein.

Section III of this dissertation was concerned with the task of explaining the identified challenges of online bio-inspiration seeking. This required a theoretical understanding of the causal processes underlying the task. In Chapter 6, such an understanding was constructed in the form of Interactive Analogical Retrieval (IAR) model, which provided a dynamics account of online bio-inspiration seeking. This leads to the next claim of this dissertation:

Claim 2: The Interactive Analogical Retrieval model provides an information-processing account of the causal mechanism underlying the task of online bio-inspiration seeking.

In Chapter 7, The IAR model was then used to provide theory-based explanations for the identified challenges, which were then used to trace the underlying causes

associated with the challenges of online bio-inspiration seeking. The identified causes leads to the third claim of this dissertation:

Claim 3: The causes for the identified challenges of online bio-inspiration seeking can be traced to certain features, or lack thereof, in current common online information environments. In particular: (3a) The issue of *low rate of encountering useful information resources issue* can be traced to the current keyword-based methods of indexing and accessing information resources in online information environments, which support access to information resources based on literal similarity (word-for-word matching) while ignoring the three pressures governing the process of analogical retrieval. (3b) The issue of *high rate of recognition errors issue* can be traced to the nature of proximal cues that one customarily encounters in current online information environments – specifically, their lack of affordance for accurately perceiving the information scent (analogical similarity) of the resources they represent. (3c) The *conceptual understanding difficulty* issue can be attributed to the fact that existing biological information resources are usually created by and for (expert) biologists, which, for non-biologists, may not contain explanations at the right level of abstraction and may leave a lot of information implicit that constituting gaps in knowledge.

Section IV of this dissertation was concerned with addressing the challenges. Once the causes were known, the next step involved proposing measures that can be implemented which ameliorates the identified challenges of online bio-inspiration seeking by specifically targeting those causes. Chapter 8 discusses the three proposed measures for ameliorating the identified challenges of online bio-inspiration seeking, which form the core of the fourth claim put forth in this dissertation:

Claim 4: The identified challenges of online bio-inspiration seeking can be ameliorated by changing the conditions in the external online information environment that make it favorable for engaging in the task of online bio-inspiration seeking. In particular: (4a) The issue of *low rate of encountering relevant information resources* can be ameliorated by semantically indexing and accessing biological information resources using concepts and relations derived from corresponding functional models. (4b) The issue of *high rate of recognition errors* can be ameliorated by enhancing proximal cues with visual overviews derived from corresponding functional models. (4c) The *conceptual understanding difficulty* issue can be ameliorated by supplying biological information resources with their corresponding functional models as external representational aids to scaffold designers' understanding.

Based on these proposed measures in Claim 4, I developed an online information-seeking environment called *Biologue* that is intended to better support the needs of designers engaged in online bio-inspiration seeking. Details about the design and implementation of Biologue can be found in Chapter 9. One unique feature of Biologue is *model-based tagging*, which represents a social approach to establishing an online corpus of biology article references annotated by their corresponding functional models. It is based on the principle of social bookmarking and is aimed to promote the sharing of biology articles among the designer community. In the same way that it is possible to catalog web pages (with Furl and del.icio.us) or photographs (with Flickr), designers can share information on biology articles in Biologue with semantic tools developed for the purpose of annotating and discovering articles using functional models. The functional models that are collectively annotated by the users are leveraged by Biologue to implement features such as model-based indexing and retrieval (first hypothesis), proximal cues enhanced by visual overviews (second hypothesis), and scaffolding for aiding designers' comprehension of biology articles (third hypothesis).

In order to validate the proposed measures, I conducted a series of evaluation studies, including three experimental studies and a pilot deployment study. Details of these studies can be found in Chapters 10 and 11, but the results are briefly summarized here.

Trends from experimental study 1 indicate that the average rate of encountering relevant information resources was significantly higher when biology documents were indexed and accessed using functional models when compared to keyword-based indexing and retrieval. Trends from experimental study 2 indicate that the average number of recognition errors was significantly lower when proximal cues were enhanced with visual overviews derived from functional models when compared to conventional proximal cues. Trends from experimental study 3 indicate that the presence of functional models in addition to textual descriptions affords deeper understanding of biological systems when compared to textual descriptions alone. In the deployment study, overall trends in self-reported data indicate that having functional models in addition to biology articles helped subjects better recognize the relevance of articles, but did not help them during retrieval or comprehension. However, among a subset of subjects who used model-based tagging feature more extensively and contributed multiple functional models in Biologue, self-reported data indicate that having models did help them during comprehension, specifically, it made reading of articles easier and go faster.

The trends from these studies are encouraging and seem to suggest that measure put forth indeed change the dynamics of online bio-inspiration seeking in favor of ameliorating the identified challenges of online bio-inspiration seeking.

Broader Implications

We can talk about the broader implications of this work along at least two dimensions: implications to other tasks and domains, and educational implications of this work.

Other tasks and domains

The present work can be generalized to other tasks and domains. The first level of generalization is going from biologically inspired design to systems biology or synthetic biology, where we are still looking at biology as a source and the currency still remains biology articles, but we are now looking at them in the context of, say analysis and synthesis. The second level of generalization is when we are not looking at biological systems, but are dealing with say, chemical, physical or systems of other kinds. Yet another kind of generalization is when we look at the inverse relationship: instead of thinking about engineers looking for biology articles, we can imagine biologists looking for engineering systems. For instance, if I am a biologist studying a particular organism, if I can find an engineering system similar to this organism for which the mathematics has been worked out, then that might provide me with analytical tools for understanding my biological system.

In all these other scenarios that also deal with systems thinking, the basic principles of my research (which is finding the right kinds of abstractions that work, and annotating article with such abstractions) can be applied in order to help practitioners. But, as we go further and further away from the current domain, less and less of the specific annotations that I have used will turn out to be useful, but the notion of semantic tagging with the right kind of abstractions will still remain. In general, one can always imagine two domains where people are likely to be experts in one domain but not the other and still need access to information from the less familiar domain. Under such circumstances, my work emphasizes gaining insights into the key domain-bridging abstractions that feature in the existing practices of practitioners (which can be obtained by careful *in situ* studies), and developing a system for semantically tagging information resources in the second domain using those abstractions.

Educational implications

The following implications can be drawn from my work from the perspective of improving the learning that occurs in the ME/ISyE/MSE/PTFe/BIOL 4803 course, which is the immediate context that defines my research.

First, helping students learn about biological knowledge from a design perspective is one of the stated learning objectives of this course. Because one of the principles of my work is to provide functional model-based annotations to biology articles, and because functional models are essentially coming from thinking about systems from a design perspective, those annotations immediately serves the purpose of that learning objective. Creating and having functional models of biology articles enables students to think about biological systems from a design perspective.

Second, if the students in this course are spending a significant amount of their time looking for biology articles on the Web in order to find their sources of inspiration, that means that they are spending a lot less time learning about other things that might be potentially more useful with respect to the objectives of the course. For example, iteration over the design process might suffer as a result of spending too much time on online bio-inspiration seeking process because there is an opportunity cost associated with it. In fact, the purpose of the course is not to learn how to do search better, and all the other skills that instructors want students to learn may not get sufficient attention. Therefore, any measures that can be taken to make the search more efficient and effective frees up time and provides more opportunities for students and instructors to focus on the actually intended learning objectives. Therefore, a tool like Biologue might be very valuable in the context of such a course.

However, a counter argument can be made that the fact that students are taking such a long time to do the search is a good thing. It means they are learning a lot of biology in the process. While perhaps that is true to some degree, the question is: the specific biological systems that students learn as part of the search, are those what they ought to be learning as part of this course? Or, are their time better spent trying to learn

design skills and other meta-knowledge required to carryout biologically inspired design more effectively? The answer is not clear, but the effective use of a tool like Biologue in the classroom will at least allow the opportunity to pose this as a problem of trade-off.

Summary of contributions

For exposition purposes this dissertation was organized as a sequence of topics suitable for individual chapters. Although this groups the contributions according to those topics (field studies, phenomenon, theory, design, evaluation, etc.), it is also possible to categorize the contributions according to what types of advances they provide for which community (cognitive science, biologically inspired design, human-information interaction, etc.). This latter method of organization is used here. Using this method of decomposition there were seven primary contributions made. Figure 13.2 depicts an incidence matrix showing how the contributions are spread out across the chapters. Short summations of these contributions follow.

Contributions to the biologically inspired design research community

First detailed field studies: Field studies of designers engaged in biologically inspired design was provided in this dissertation. To my knowledge, this represents the first detailed field studies of biologically inspired design practice. These studies also identified the actual issues that confront designers in the field, which have been largely overlooked by other similar research endeavors. These issues provide the basis for theorizing about what sort of support needs to be provided to designers during the bio-inspiration seeking process.

Theory of bio-inspiration seeking: Throughout this work, the guiding principle was that a solid theory-based understanding of bio-inspiration seeking support is possible, and the time has come for BID research to begin developing and using theories in earnest. As of now craft knowledge and folk psychology fill the theoretical void. This dissertation

not only proposes a theory of bio-inspiration seeking, but also offers a methodology for technology development that is grounded in the proposed theory.

Biologue: Biologue is the culmination of the research into bio-inspiration seeking phenomenon in the form of a tool to support this phenomenon. The design principles behind Biologue, if not the tool itself in its present form, are a valuable contribution to the BID community. Biologue represents a well-thought-out and a well-argued-out blueprint for building a tool to support the process of online bio-inspiration seeking.

Contributions to the human-information interaction research community

Extensions to Information foraging theory: An extension was proposed to the current models of information foraging theory that take into account the peculiarities of information seeking for the purposes of analogy making.

New information scent model: A new information scent model called PRISM (Pressurized Information Scent Perception Model) was also developed here. This provides an extension to the information scent model which was part of the original Information Foraging Theory framework.

Model-based tagging: Keyword-based tagging, to a larger extent, and semantic tagging, to a lesser extent, are commonly researched interaction techniques with the HII community for the purposes of indexing and organizing online resources in a social fashion. Model-based tagging, although an extension of these interaction techniques, represents an innovation in this sphere. Model-based tagging not only serves the primary purpose of (conceptually) indexing, organizing, and accessing online resources, but also the secondary purposes of: (1) aiding the recognition of those information resources, and (2) scaffolding the comprehension of those information resources.

Contributions to the cognitive science community

Interactive analogical retrieval model: This account provides a new perspective on analogical retrieval, a perspective that has not received sufficient attention in the past. While most theories of analogical retrieval are oriented towards explaining retrieval of source analogues from the long-term memory of individual agents, interactive analogical retrieval account explains how agents obtain source analogues through interaction with the external information environment when those source analogues are not *a priori* encoded in the long-term memory of the agent.

Future Work

Looking forward, I see several opportunities for further research in these areas under four genres: (1) expanding demography; (2) quality of emergent models; (3) scalability; and (4) HCI for model-based tagging.

Expanding demographics

Most of my research has targeted the BID student population. This includes everything from the *in situ* studies to the deployment of Biologue. One of the ways in which this research can be furthered is by expanding the demographics to include practicing biologically inspired design professionals. This includes expanding the research to include research labs and industrial settings where "real" biologically inspired design takes place. I assume that the needs of professional will be slightly if not significantly different from the needs of students. The next *in situ* study that I would conduct would be in a professional setting. Likewise, I would take the ideas that were developed in the context of classroom and share it with the professional community and get their feedback. I would also share Biologue with this community and try to establish an online "living" community of practitioners around Biologue. It is clearly important to vet the ideas developed in this research through its use in multiple scenarios and multiple communities.

Scalability

The success of the solution approach presented in this work beyond the laboratory conditions relies on the establishment of a large-scale information-seeking environment that has certain qualities laid out in my hypotheses. This immediately raises the issue of scalability. How can we establish an environment on the scale of, say Google Scholar, such that the articles in them are tagged with SBF models? This research question although very pertinent, currently lies beyond the current scope of my research. My research only claims that *if* we have such an environment at our disposal, then we can expect to see certain benefits with respect to the efficiency of the online bio-inspiration seeking process.

Currently, there are no straightforward ways for achieving such an environment though automated means, unless information extraction algorithms mature to such an extent that SBF models can be extracted directly from natural language texts. In the absence of such automated means, one way to achieve scale is by relying on mass participation and social contribution. That is the basic idea behind the model-based tagging feature in Biologue. But whether this technique can achieve the scale that is required remains to be seen. More generally, the notion of collective knowledge systems relies on social action to obtained structured information, which is then fed back to provide the kind of services that are deemed useful by the contributors. There are interesting theoretical and empirical research questions surrounding collective knowledge systems that are directly relevant to furthering the research presented here.

HCI for model-based tagging

Model-based tagging is a concept that can be implemented in several ways in a practical application. There are several interesting HCI issues surrounding the implementation of this concept. The current implementation of model-based tagging is pretty rudimentary. This may or may not be the best approach to implementing such a

concept. An alternative way, for instance, is to implement it as a hash tag system (as in Twitter), but the hash tags are richer and contain embedded structure. Since hash tags are more familiar to the users, the uptake of this idea might be higher compared to the current implementation. More generally, the research question of interaction techniques to foster the elicitation of structured information is another fruitful research area to pursue as an extension to the current research.

The nature and quality of emergent models

Let us assume that the scalability and HCI issues are tackled. Let us also assume that we have a large online community of people citing articles and tagging them on a regular basis. In such a scenario, we would expect to see the emergence of (partial) SBF models of biological systems associated with the articles that are cited. But what can we say about the quality of these emergent models? Because these models are created in a distributed fashion, these models may be incomplete, incorrect, and/or mutually conflicting. What does the quality of models even mean in this context? How do we ensure that the quality of these models meet minimum standards? What checks and balances do we put in place to reduce the proliferation of low quality of models? How can we trust that the models are consistent with the article contents? What happens when two people tag the same article with mutually incompatible models? These are all valid questions that one can ask of my research for which I do not have good answers at this stage. This opens up new directions of research in the future.

Application of natural language processing and machine learning techniques

Model-based tagging is a user-contributed manual approach to annotating the biology articles with functional model. But one can imagine automated techniques for accomplishing the same. Information extraction and other NLP techniques can be used to extract structured information from unstructured textual documents. In principle, one can

apply such techniques in order to extract functional models (or at least partial models) automatically from biology articles. How well do these techniques work in this context is an interesting research question to be investigated.

Similarly one can imagine supervised machine learning techniques to accomplish the same. If we have a corpus of biology articles already annotated with functional models in Biologue, these can become the training set for training machine learning algorithms to generate functional models of the rest of the available set of biology articles.

The automated techniques opens up new possibilities for scaling the process of annotation, but at the same time raises a number of interesting issues related to the quality and the reliability of the automatically created annotations.

APPENDIX A

The coding scheme that was used for coding video data in the Find-frequency study discussed in Chapter 10 titled "Experimental Studies".

Code		**Sub-code**	**Type**	**Comments**
ISC	Initiate search (control)	SS	String	Search string
		#SR	Integer	Number of search results
		#DA	Integer	Number of designated articles
ISE	Initiate search (Experiment)	FLD	Enumeration	Field used for search
		SS	String	Search string
		?AUTO	Boolean	Auto suggested?
		#SR	Integer	Number of search results
		#DA		Number of designated articles
ESR	Eyeball search results	T	Integer	Time (seconds)
APC	Attend to proximal cues	T	Integer	Time (seconds)
		?DES	Boolean	Proximal cues belong to designated article or not
		?MOD	Boolean	Proximal cues contain SBF model or not
		TITLE	String	Title of the article
ABIB	Attend to bibliographic information in the proximal cues	T	Integer	Time (seconds)

AF	Attend to Function	T	Integer	Time (seconds)
		FANME	String	Function name
ASB	Attend to SB information	T	Integer	Time (seconds)
		FNAME	String	Function name associated with SB
AIP	Attend to information patch (article)	T	Integer	Time (seconds)
		TITLE	String	Title of the article
		SS	Integer	Strength of information scent (self-reported) (1-5)
ATH	Attend to thesaurus	T	Integer	Time (seconds)
		TERM	String	Term being looked up
PI	Positive identification	TITLE	String	Title of the article
		RAT	String	Rationale

REFERENCES

Agnar, A., & Plaza, E. (1994). *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*, Artificial Intelligence Communications, 7: 39-52.

Abel F., Frank M., Henze N., Krause D., Plappert D. & Siehndel P. (2007). Groupme! - Where Semantic Web meets Web 2.0. In ISWC/ASWC, volume 4825 of LNCS, p. 871–878: Springer.

Agichtein, E., & Gravano, L. (2000). Extracting relations from large plain - text collections, in: Proceedings of the 5th ACM International Conference on Digital Libraries (DL-00), San Antonio, Texas, 2000, pp. 85–94.

Aizenberg, J., Tkachenko, A., Weiner, S., Addadi, L. & Hendler, G. (2001). Calcitic microlenses as part of the photoreceptor system in brittlestars. *Nature*, 412, 819-822.

Altshuller, G. (1984). *Creativity as an exact science*. Gordon and Branch Publishers, Luxembourg.

Angeletou, S. (2008). Semantic Enrichment of Folksonomy Tagspaces. International Semantic Web Conference 2008: 889-894

Arzt, E., Gorb, S. & Spolenak, R. (2003) From micro to nano contacts in biological attachment devices. *Proceedings of the National Academy of sciences*, 100(19), 10603-10606.

Ashley, S. (2003). Artificial Muscles. *Scientific American*, 289(4), 53-59.

Ashley, S. (2004). Bumpy flying. Scalloped flippers of whales could reshape wings. *Scientific American*, 2004 Aug; 291(2):18, 20.

AskNature.org. (2012). "Ask Nature: A Project of the Biomimicry Institute." Accessed September 1. http://www.asknature.org/.

Auer, S. & Lehmann, J. (2007). What have Innsbruck and Leipzig in common? Extracting semantics from Wiki content, in Fourth European Semantic Web Conference (ESWC 2007), June 3–7, 2007.

Bar-Cohen, Y. (2005). *Biomimetics - Biologically Inspired Technologies*, Boca Raton, FL: CRC Press, pp 1–552

Baurelein, E. (2003). Biomineralization of Unicellular Organisms: An Unusual Membrane Biochemistry for the Production of Inorganic Nano- and Microstructures. *Angew. Chem. Int. Edn.*, 42(6), 614- 641.

Benyus, J. (1997). *Biomimicry: Innovation Inspired by Nature*. New York: William Morrow.

Biomimicry Institute (2009). Biomimicry: A Tool for Innovation. http://www.biomimicryinstitute.org/about-us/biomimicry-a-tool-for-innovation.html. Last retrieved on April 25, 2011.

Bizer, C., Cyganiak, R., & Gauß, T. (2007). The RDF Book Mashup: from Web APIs to a Web of data, in: Third Workshop on Scripting for the Semantic Web, ESWC, Innsbruck, Austria, June 6, 2007

Boehm, B. W. (1988). A Spiral Model of Software Development and Enhancement, *IEEE Computer*, 21(5), pp. 61-72.

Bosner, R. & Vincent, J. (2006). Technology trajectories, innovation, and the growth of biomimetics, *Journal of Mechanical Engineering Science*, 221, pp 1177-1180

Breslin, J. G., Passant, A., & Decker, S. (2009). *The Social Semantic Web*, Springer-Verlag: Berlin Heidelberg

Brott, L. L., Naik, R. R., Pikas, D. J., Kirkpatrick, S. M., Tomlin, D. W., Whitlock, P. W., Clarson, S. J., & Stone, M.O. (2001). Ultrafast holographic nanopatterning of biocatalytically formed silica. *Nature*, 413, 291-293.

Cafarella, M., Downey, D., Soderland, S., & Etzioni, O. (2005) KnowItNow: fast, scalable information extraction from the Web, in: Proceedings of the Human Language Technology Conference (HLT-EMNLP-05), Vancouver, Canada, 2005, pp. 563–570.

Card, S., Pirolli, P., Van Der Wage, M., Morrison, J., Reeder, R. W., Schraedley, P. K., & Boshart, J. (2001). Information scent as a driver of web behaviour graphs: Results of a protocol analysis method for Web usability. in: Proceedings of CHI 2001, 498-505.

Chakrabarti, A., Sarkar, P., Leelavathamma, B., & Nataraju, B.S. (2005). A functional representation for aiding biomimetic and artificial inspiration of new ideas. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 19, 113–132.

Collins, S., Ruina, A., Tedrake, R., & Wisse, M. (2005). Efficient Bipedal Robots Based on Passive- Dynamic Walkers. *Science*, 307, 1082-1085.

Cross, N. (1982). Designerly ways of knowing, *Design Studies*, vol. 3, pp. 221-227.

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach, *BMC Medical Research Methodology*, 11:100.

Dabbs, D.M., & Aksay, I.A. (2000). Self-assembled ceramics produced by complex-fluid templation, *Annu. Rev. Phys. Chem.*, 51, 601-622.

Darden, L. (1983). Artificial intelligence and philosophy of science: Reasoning by analogy in theory construction. In P. Asquith & T. Nickles (Eds.), *PSA* 1982 (Vol. 2). East Lansing, MI: Philosophy of Science Society.

Darden, L. (1991). *Theory Change in Science: Strategies from Mendelian Genetics*. New York: Oxford University Press.

Darden, L. (1996). Generalizations in Biology, *Studies in the History and Philosophy of Science*, 27:409-419.

Dertouzos, M. (2001). *The Unfinished Revolution: Human-centered Computers and What They Can Do for Us*. HarperCollins: New York NY.

Dickinson, M.H. (1999) Bionics: Biological insight into mechanical design, *Proceedings of the National Academy of sciences*, 96(25), 14208- 14209.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1998). *Human-Computer Interaction*, Prentice Hall: New Jersey.

Dunbar K. (1995). How scientists really reason: scientific reasoning in real-world laboratories. In: Sternberg R, Davidson J, editors. *The Nature of Insight*. Cambridge, MA: MIT Press; 1995. pp. 365–96.

Dunbar, K. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings yet so difficult in the psychological laboratory. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 313-334). Cambridge, MA: MIT Press.

Dunbar, K. & Blanchette, I. (2001). The invivo/invitro approach to cognition: the case of analogy. *Trends in Cognitive Sciences*, 5, 334--339.

Eisenhardt, K. M. (1989). Building theories from case study research, *Academy of Management Review*, 14: 488- 511.

Falkenhainer, B., Forbus, K. D. & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples, *Artificial Intelligence*, 41(1): 1–63.

Farrell, J.A., Pang, S., & Li, W. (2005). Chemical Plume Tracing via an Autonomous Underwater Vehicle, *IEEE Journal of Oceanic Engineering*, 30(2), 428-442.

Fish, F.E. (2006). Limits of nature and advances of technology: What does biomimetics have to offer aquatic robots, *Applied Bionics and Biomechanics*, 3(1), 49-60.

Flanagan, J., Huang, T., Jones, P., & Kasif, S. (1997). *Human-centered systems: Information, interactivity, and intelligence*, Technical report, National Science Foundation.

Flynn, E. C., Lee, S. W., Peelle, B. R., & Belcher, A. M. (2003). Viruses as vehicles for growth, organization and assembly of materials. *Acta Materialia*, 51, 5868 – 5880.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC:A model of similarity-based retrieval. *Cognitive Science*, 19, 141-205.

Fountopoulos, G. (2007). *RichTags: A Social Semantic Tagging System*, Dissertation for MSc Web Technology, University of Southampton, Southampton, December 2007.

French, M. (1994). *Invention and evolution: design in nature and engineering*. 2nd edition. Cambridge University Press, 1988.

Full, R. J. & Koditschek, D. E. (1999). Templates and Anchors: Neuromechanical Hypotheses of legged locomotion on land. *The Journal of Experimental Biology*, 202, 3325–3332.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.

Goel, A. K. (1997). Design, analogy and creativity, *IEEE Expert Intelligent Systems and Their Applications*, 12(3), 62–70.

Goel, A. K., Rugaber, S., & Vattam, S. (2009). Structure, behavior & function of complex systems: The SBF modeling language. *International Journal of AI in Engineering Design, Analysis and Manufacturing*, 23, 23–35.

Golder, S. & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems, *Journal of Information Science*, 32(2), 198 - 208

Gray, B. (2006). Artificial Arrays Could Help Submarines Make Like a Fish, *Science*, 313, 1382-1383.

Giere, R. (2002). Models as parts of distributed cognitive systems. In L. Magnani & N. Nersessian (Eds.), *Model based reasoning: Science, technology, values*. Kluwer Academic Publishers: Dordrecht.

Giere, R. (2004a). How Models Are Used to Represent Reality, *Philosophy of Science* 71, *Supplement*, S742-752.

Giere R. (2004b). The problem of agency in scientific distributed systems. *Journal of Cognition and Culture*, 4(3):759–774.

Gooding, D. C. (2004). Cognition, construction and culture: Visual theories in the sciences. Journal of Cognition and Culture, 4, 551–594.

Halpin, H., Robu, V. & Shepard, H. (2006). The dynamics and semantics of collaborative tagging, in: Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW'06), 2006.

Halverson, C. A. (2002). Activity theory and distributed cognition: Or what does CSCW need to DO with theories? *Computer Supported Cooperative Work*, 11: 243–267.

Haven, K. (2006). *100 Greatest Science Inventions of All Time*, Libraries Unlimited: Westport, CT.

Helms, M., Vattam, S., & Goel, A. (2009). Biologically Inspired Design: Process and Products, *Design Studies*, 30(5), 606-622.

J. Hunter. (2009). Collaborative Semantic Tagging and Annotation Systems, *Annual Review of Information Science and Technology*, volume 43, chapter 2.

Hoffman, R. R., Hayes, P. J., & Ford, K. M. (2001). Human-Centered Computing: Thinking In and Outside the Box, *IEEE Intelligent Systems*, 16(5), 76–78.

Hoffman, R. R., Feltovich, P., Ford, K. M., Woods, D., Klein, G., & Feltovich, A. (2002). A Rose by Any Other Name ... Would Probably Be Given an Acronym, *IEEE Intelligent Systems*, 17(4), 72 - 80.

Hoffman, R. R. (2004). *Human-centered computing principles for advanced decision architectures*, Technical Report, Army Research Laboratory.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295-355.

Holyoak, K. & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, 15, 332-340.

Holyoak, K. & Thagard, P. (1995). *Mental Leaps: Analogy in Creative Thought*. MIT Press, Cambridge, MA.

Hutchins, E. (1996). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Hutchins, E. & Klausen, T. (1998). Distributed cognition in an airline cockpit, in *Cognition and Communication at Work*, Y. Engestrom and T. Middleton, Eds. Cambridge, U.K.: Cambridge Univ. Press, 15–34.

Jeong, K. H., Kim, J., & Lee, L.P. (2006). Biologically Inspired Compound Artificial Eye, *Science*, 312, 557-561.

Kim H. L., Yang S. K., Song S. J., Breslin J. G. & Kim H. G. (2007). Tag Mediated Society with SCOT Ontology. In Semantic Web Challenge, ISWC 2007.

Kling, R. & Star, L. (1998). Human Centered Systems in the Perspective of Organizational and Social Informatics, *Computers and Society*, 28(1), 22–29.

Kolodner, J. L. (1993). *Case-based reasoning*. Kaufmann: San Mateo, CA.

Kolodner, J. L. (1994). From natural language understanding to case-based reasoning and beyond: a perspective on the cognitive model that ties it all together. In Schank,

R. C. and Langer, E. (eds) *Beliefs, Reasoning, and Decision Making: Psycho-Logic in Honor of Bob Abelson*. Lawrence Erlbaum: Hillsdale, NJ.

Kolodner, J. L. (1997). Educational implications of analogy. *American Psychologist*, 52, 57–66.

Kruk, S. R., Woroniecki, T., Gzella, A., Dabrowski, M., & McDaniel, B. (2007). Anatomy of a social semantic library, in: European Semantic Web Conference, Volume Semantic Digital Library Tutorial 2007.

Kruk, S. R., Samp, K., O'Nuallain, C., Davis, B., McDaniel, B., & Grzonkowski, S. (2006). Search interface based on natural language query templates. In Proceedings of the poster session of IADIS International Conference WWW/Internet 2006.

D. Lin, & Pantel, P. (2002). Concept discovery from tex, in: Proceedings of the 19th International Conference on Computational Linguistics (COLING-02), Taipei, Taiwan, 2002, pp. 1–7.

Lita, L., & Carbonell, J. (2004). Instance-based question answering: a data driven approach, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-04), Barcelona, Spain, 2004, pp. 396–403.

Leake, D. (1996). CBR in Context: The Present and Future, In Leake, D., editor, *Case-Based Reasoning: Experiences, Lessons, and Future Directions*. AAAI Press/MIT Press, 1996, 1-30.

Leake, D. (1998). Case-Based Reasoning. In W. Bechtel & G. Graham (eds.), *A Companion to Cognitive Science*. Blackwell.

Lee L. P., & Szema, R. (2005). Inspirations from Biological Optics for Advanced Photonic Systems, *Science*, 310, 1148 – 1150.

Linsey, J. S., Murphy, J. T., Markman, A. B., Wood, K. L., & Kurtogula, T. (2006). Representing Analogies: Increasing the Probability of Innovation, in: the Proceedings of IDETC/CIE, ASME, Philadelphia, PA.

Linsey, J., Wood, K., & Markman, A. (2008), Modality and Representation in Analogy, *Artificial Intelligence for Engineering Design and Manufacturing*, 22(2), 85 - 100.

Linsey, J., Markman, A., & Wood, K. (2008). WordTrees: A Method for Design-by-Analogy, Proceedings of the 2007 American Society for Engineering Education Annual Conference Pittsburg, PA.

von Ahn, L. (2006). Games with a purpose, *IEEE Computer Magazine*, 96–98.

Mak, T., & Shu, L. (2008). Using descriptions of biological phenomena for idea generation, *Research in Engineering Design*, 19(1), 21 - 28.

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). Position paper, tagging, taxonomy, flickr, article, toread, In *Collaborative Web Tagging Workshop at WWW2006*, Edinburgh, UK.

Michelson, R. C. (2004) Novel Approaches to Miniature Flight Platforms, In *Proceedings of the Institute of Mechanical Engineers*, Part G: Journal of Engineering, 218(6), 363-373.

Mika, P. (2005). Ontologies are us: A unified model of social networks and semantics, In Y. Gil et al. (Eds.) *Proceedings of the 4th International Semantic Web Conference (ISWC 2005)*, Galway, Ireland, November 6-10, 2005, pp. 522–536. Berlin: Springer-Verlag.

Miller, D. & Brickley, D. (2011) *Friend of a Friend Project*, http://xmlns.com/foaf/spec/, Setember 2011

Miles, A. & Brickley, D. (2005). SKOS Core Vocabulary Specification, W3C Working Draft, http://www.w3.org/TR/swbp-skos-core-spec (current 10 May 2005).

Mooney, R. & Bunescu, R. (2005). Mining knowledge from text using information extraction, *SIGKDD Explore*, 7 (1) (2005) 3–10.

Morse, T. M., Ferree, T. C., & Lockery, S. R. (1998). Robust spatial navigation in a robot inspired by chemotaxis in C.elegans, *Adaptive Behavior*, 6, 393-410.

Nagel, R., Tinsley, A., Midha, P., McAdams, D., Stone, R., & Shu, L. (2008). Exploring the use of functional models in biomimetic design, *Journal of Mechanical Design* 130(12), 11–23.

Nagel, J., Stone, R., & McAdams, D. (2010). An engineering-to-biology thesaurus for engineering design, in: *Proceedings of ASME 2010 IDETC/ CIE Conference*, 2010, Montreal.

Nardi, B. (1992). Studying context: A comparison of activity theory, situated action models and distributed cognition, in: *Proceedings East-West HCI Conference,* pp. 352–359, St. Petersburg, Russia, August 4–8.

Nersessian, N. (1985). Faraday's Field Concept. In *Faraday Rediscovered*, ed. D. Gooding and F. James, pp. 175-87. London: Macmillan

Nersessian, N. (1986). A Cognitive-Historical Approach to Meaning in Scientific Theories. In *The Process of Science: Contemporary Philosophical Approaches to Understanding Scientific Practice*, ed. N. J. Nersessian, pp. 161-78. Dordrecht: Martinus Nijhoff.

Nersessian, N. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. Giere (Ed.), *Cognitive models of science* (pp. 3-44). Minneapolis: Universityof Minnesota Press.

Nersessian, N. (2002). The cognitive basis of model-based reasoning in science. In Carruthers, P., Stich, S. & Siegal, M. (eds.) *The Cognitive Basis of Science*, 133 - 153. Cambridge University Press.

Nersessian, N. J., Kurz-Milcke, E., Newstetter, W. C., & Davies, J. (2003). Research laboratories as evolving distributed cognitive systems. In: *Proceedings of The 25th Annual Conference of the Cognitive Science Society*. pp. 857 - 862.

Nersessian, N. J. (2005). Interpreting Scientific and Engineering Practices: Integrating the Cognitive, Social and Cultural Dimensions. Chapter in *Scientific and Technological Thinking*. Earlbaum. 17-56.

Nersessian, N. J. (2006). The cognitive-cultural systems of the research laboratory. *Organization Studies*, 27, 125–145.

Nersessian, N. (2008). *Creating scientific concepts*. Cambridge, MA: MIT Press.

O'Reilly, T. (2011). *What is Web 2.0: Design patterns and business models for the next generation of software*, September 2011, http://oreilly.com/web2/archive/what-is-web-20.html

Pantel, P., Ravichandra, D. (2004). Automatically labeling semantic classes, in: *Proceedings of the 2004 Human Language Technology Conference (HLT- NAA Cl-04)*, Boston, MA, 2004, pp. 321–328.

Passant, A. (2007). Using ontologies to strengthen folksonomies and enrich information retrieval in weblogs. In *Proceedings of the First International Conference on Weblogs and Social Media (ICWSM)*, Boulder, Colorado, 2007.

Popper, K., Merson R. L., & Camirand W. M. (1968). Desalination by osmosis-reverse osmosis couple, *Science*, Mar 22; 159 (821), 1364 - 1365.

Reilly, R. C. (2010). Participatory case study. In: *Encyclopedia of case study research*. Sage, Thousand Oaks, CA, pp. 658-660.

Riesbeck, C., & Schank, R. (1989). *Inside Case-based Reasoning*. Northvale, NJ: Erlbaum, 1989.

Sarikaya, M., Tamerler, C., Jen, A.K.-Y., Schulten, K., & Baney, F. (2003). Molecular biomemetics: nanotechnology through biology. *Nature Materials*, 2(9), 577-585.

Sarkar, P., Phaneendra, S., & Chakrabarti, A. (2008). Developing Engineering Products Using Inspiration from Nature, *Journal of Computing and Information Science in Engineering*, 8(3), pp. 1-9.

Sartori, J., Pal, U., & Chakrabarti, A. (2010). A methodology for supporting "transfer" in biomimetic design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 24, 483-505.

Scott, R., Roth, E. M., Deutsch, S. E., Malchiodi, E., Kazmierczak, T. E., & Eggleston, R. G. (2005). Work-Centered Support Systems: A Human-Centered Approach to Intelligent System Design, *IEEE Intelligent Systems*, March/April 2005.

Seifert, C. (1989). Analogy and case-based reasoning. In: *Proceedings: Case-Based Reasoning Workshop*. Morgan Kaufmann, San Mateo, CA.

Singh, V., Skiles, S. M., Krager, J. E., Wood, K. L., Jensen, D., & Szmerekovsky, A. (2006). Innovations in Design through Transformation: A Fundamental Study of Transformation Principles, in: *Proceedings of IDETC/CIE 2006*, ASME: Philadelphia, PA

Srinivasarao, M. (1999). Nano-optics in the Biological World: Beetles, Butterflies, Birds and Moths. *Chemical Reviews*, 99(7), 1935-1961.

Stone, R., & Wood, K. (2000). Development of a functional basis for design. *Journal of Mechanical Design*, 122(4), 359-370.

Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence*, 46, 259-310.

Tinsley, A., Midha, P., Nagel, R., McAdams, D., Stone, R., & Shu, L. (2007). Exploring the use of functional models as a foundation for biomimetic conceptual design. *ASME Design Theory and Methodology Conference*, Las Vegas, NV, DETC2007-35604.

Tweney, R. D. (1989). A framework for the cognitive psychology of science. In B. Gholson, W. R. Shadish Jr, R. A. Neimeyer, & A. C. Houts (Eds.), *Psychology of science: Contributions to metascience* (pp. 342–366). New York: Cambridge University Press.

Vakili, V., Chiu, I., Shu, L. H., McAdams, D., Stone R. (2007). Including Functional Models of Biological Phenomena as Design Stimuli, in *ASME International*

Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Las Vegas, USA. 2007.

Vattam, S., Helms, M., & Goel, A. K. (2010). Biologically Inspired Design: A Macrocognitive Account. In *Proceedings of ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2010)*. August 15-18, 2010, Montreal, Quebec, Canada.

Vattam, S., Wiltgen, B., Helms, M., Goel, A., & Yen, J. (2010). DANE: fostering creativity in and through biologically inspired design. In: *Proceedings of 1st Int. Conf. Design Creativity*, Kobe, Japan, November 2010.

Vincent, J., & Mann, D. (2002) Systematic Transfer from Biology to Engineering. *Philosophical Transactions of the Royal Society of London*, 360: 159- 173.

Vincent, J., Bogatyreva, O., Bogatyrev, N., Bowyer, A., & Pahl, A-K. (2006). Biomimetics - its practice and theory. *Journal of the Royal Society Interface*, 3(9), 471 – 482.

Vogel S. (1998). *Cats' Paws and Catapults: Mechanical Worlds of Nature and People*. New York: Norton

Vukusic, P. & Sambles, J.R. (2003) Photonic structures in biology. *Nature*, 422, 852-855.

Walenstein, A. (2002). *Cognitive Support in Software Engineering Tools: A Distributed Cognition Framework*, doctoral dissertation, Computing Science Dept., Simon Fraser Univ., Burnaby, B.C., Canada, 2002.

Wiltgen, B., Goel, A. & Vattam, S. (2011). Representation, Indexing, and Retrieval of Biological Cases for Biologically Inspired Design. In: *Proceedings of 19th International Conference on Case Based Reasoning (ICCBR)*, Greenwich, London, September 2011.

Winograd, T., & Woods, D. D. (1997). Challenges for Human-Centered Design. In: J. Flanagan, T. Huang, P. Jones & S. Kasif (Eds), *Human-Centered Systems:*

Information, Interactivity, and Intelligence. Washington, D.C.: National Science Foundation, July.

Yen, J. & M.J. Weissburg. (2007). Perspectives on biologically inspired design: introduction to the collected contributions. *Bioinspiration & Biomimicry*. 2: DOI 10.1088/1748-3190/2/4/E01

Yin, R. K. (2009). *Case Study Research, Design and Methods*, Sage Publications, Thousand Oaks, CA, 2009.